REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comment regarding this burden estimates or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE January 16, 1998	3. REPORT TYPE AND DATES COVERED Final, 1 Jul 94 – 30 Nov 97

4. TITLE AND SUBTITLE Information Theory Analysis for Data Fusion	5. FUNDING NUMBERS DAAH04-94-C-0011

6. AUTHOR(S)

Ronald P. S. Mahler

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(ES) Lockheed Martin Tactical Defense Systems - Eagan 3333 Pilot Knob Road M/S U1P28 Eagan, MN 55121	8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211	10. SPONSORING / MONITORING AGENCY REPORT NUMBER ARO 32628.12-EL

11. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution unlimited.	12 b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Aims: Complete theoretical development of random set unified data fusion, and place under theoretically solid information theory foundation suitable for publication. **Findings:** (1) Showed that optimal sensor allocation (redirection of the reallocatable sensors in a sensor suite) can be subsumed within the random set approach to information fusion, via generalization of nonlinear optimal control theory. (2) Showed that random set theory and information theory provides a common basis for performance evaluation in information fusion. Showed that parameters (e.g. target-I.D. performance) can be measured in terms of information, and likewise for *user-defined constraints* (e.g. subjective or multiple definitions of information). (3) Showed that both precise and ambiguous observations can be fused by generalizing Bayesian measurement models and the standard Bayesian recursive nonlinear filtering equations. (4) Seven chapters were completed and submitted for a book   published in 1997 by Kluwer. (5) Organized a joint ONR/ARO scientific workshop on *Applications and Theory of Random Sets*.

14. SUBJECT TERMS Information fusion, Information Theory, Random sets	15. NUMBER IF PAGES
	16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UL

INFORMATION THEORY ANALYSIS FOR DATA FUSION

FINAL REPORT

AUTHOR: RONALD P.S. MAHLER

JANUARY 16, 1998   **19980519 199**

INSTITUTION: LOCKHEED MARTIN TACTICAL DEFENSE SYSTEMS
3333 Pilot Knob Road
Eagan, Minnesota 55121

1. FOREWARD

INFORMATION THEORY ANALYSIS FOR DATA FUSION

Contract DAAH04-96-R-BAA1

November 30, 1997
Final Report submitted to U.S. Army Research Office, Electronics Division, by
Ronald P.S. Mahler, Ph.D., for
Lockheed Martin Tactical Defense Systems, 3333 Pilot Knob Rd., Eagan MN 55121

This final report is offered by Lockheed Martin Tactical Defense Systems, Eagan MN, in compliance with the reporting requirements for research conducted during the last three years under contract DAAH04-94-C-0011. Under this contract Lockheed Martin has developed a unified, information theory-based approach to information fusion. The proposed theoretical tool is "finite set statistics," a special case of *random set theory* specifically developed as a part of the project. Finite-set statistics results in a systematic, fully probabilistic, theoretical unification of: detection, classification, tracking, decision-making, sensor management, expert systems theory, and performance evaluation. Highlights of this work are:

(1) Algorithms for simultaneous optimal estimation of numbers, identities, geokinematics of targets, along with control of sensor dwells and modes;
(2) Rigorous basis for fusion of "ambiguous" (imprecise, vague, contingent) observations;
(3) Systematic, theoretically justifiable performance evaluation using information theory;
(4) Optimal sensor management via nonlinear control theory and information theory.

In addition, preliminary work was begin towards developing a multisource, multitarget decision theory potentially applicable to Levels 2 fusion (situation assessment) and Level 3 fusion (threat assessment). Seventeen conference papers resulted from this work. The main result of the project, however, was a book: *Mathematics of Data Fusion* (Kluwer Academic Publishers), co-authored with I.R. Goodman of Naval Research and Development and Prof. H.T. Nguyen of New Mexico State University (Las Cruces). Chapter 2 and Chapters 4 through 8 of this book were devoted to work completed under this contract. In a closely related activity, the P.I. for this project was principal organizer, co-chair, and co-editor for a Workshop on Applications and Theory of Random Sets, held at the Institute for Mathematics and Its Applications (Minneapolis), jointly sponsored by Office of Naval Research, the Electronics Division of U.S. Army Research Office, and Lockheed Martin Tactical Defense Systems. The proceedings of this workshop appeared in hardcover (Springer-Verlag) in November 1997.

2. TABLE OF CONTENTS

Information Theory Analysis for Data Fusion (ITADF)

Contract DAAH04-96-R-BAA1

November 30, 1997
Final Report submitted to U.S. Army Research Office, Electronics Division, by
Ronald P.S. Mahler, Ph.D., for
Lockheed Martin Tactical Defense Systems, 3333 Pilot Knob Rd., Eagan MN 55121

Section/Page

Number

3. LIST OF APPENDICES, ILLUSTRATIONS AND TABLES

4. FINAL REPORT

Information Theory Analysis for Data Fusion (ITADF)

Contract DAAH04-96-R-BAA1

November 30, 1997
Final Report submitted to U.S. Army Research Office, Electronics Division, by
Ronald P.S. Mahler, Ph.D., for
Lockheed Martin Tactical Defense Systems, 3333 Pilot Knob Rd., Eagan MN 55121

This final report is offered by Lockheed Martin Tactical Defense Systems, Eagan MN, in compliance with reporting requirements for work completed under the three-year contract DAAH04-94-C-0011. This contract expired on Nov. 30, 1997.

4-A. STATEMENT OF THE PROBLEM STUDIED

4-A-a. Summary of the Problem Studied. Under contract DAAH04-94-C-0011 Lockheed Martin developed a unified, information theory-based approach to information fusion. The proposed theoretical tool is "finite set statistics" (a special case of *random set theory* specifically developed as a part of the project). Finite-set statistics is a unified statistical calculus which (1) allows multisensor, multitarget information fusion problems to be treated mathematically in the same way as single-sensor, single-target problems; and (2) provides a probabilistic framework for integrating expert systems approaches (e.g. fuzzy logic, imprecise evidence, rule-based inference). The result is a systematic theoretical unification of: detection, classification, tracking, decision-making, sensor allocation, expert systems theory, and performance evaluation. Highlights of this work are:

(1) Algorithms for optimal simultaneous estimation of numbers, identities, geokinematics of targets, along with the optimal control of sensor dwells and modes;
(2) Rigorous basis for fusion of "ambiguous" (imprecise, vague, contingent) observations;
(3) Systematic, theoretically justifiable performance evaluation using information theory;
(4) Optimal sensor management via nonlinear control theory and information theory.

In addition, under this contract preliminary work was begin towards developing a multisource, multitarget decision theory potentially applicable to Levels 2 fusion (situation assessment) and Level 3 fusion (threat assessment).

Items (1), (2) and (3) above have been described in detail in Chapters 2,4,6,7,8 of *Mathematics of Data Fusion* [12], a book published by Kluwer Academic Publishers in September 1997 (co-authors: I.R. Goodman, NRaD Code 4221, and Prof. H.T. Nguyen, New Mexico State University-Las Cruces.) Seventeen conference papers also were produced as the result of this work completed under this contract. In a closely related activity, the P.I. for this project was

principal organizer, co-chair, and co-editor for a Workshop on Applications and Theory of Random Sets, held at the Institute for Mathematics and Its Applications (Minneapolis), jointly sponsored by ONR, ARO, and Lockheed Martin. The proceedings of this workshop appeared in hardcover (Springer-Verlag) in November 1997 **[14]**.

This work has attracted considerable favorable attention and interest in both the DoD information fusion and academic communities. The P.I. was invited to speak on the subject at several DoD and university workshops and seminars, including the Air Force Institute of Technology, Naval Research and Development, the USAF Tracking and Correlation Symposium, Harvard University Department of Applied Sciences, the Johns Hopkins Department of Electronic and Computer Engineering, the University of Massachusetts (Amherst) Department of Electrical Engineering, and the New Mexico State University (Las Cruces) Department of Mathematical Sciences. He has also been invited to speak on the same subject at several engineering conferences, including the 1995 IEEE Conference on Decision and Control and the 1994 National Symposium on Sensor Fusion (formal invitations) and the 1995 and 1998 SPIE AeroSense Conferences (informal invitations).

4-A-b. The Need for This Work: Technical and Scientific. The approach investigated in this work is partially motivated by the fact that many kinds of data are *ambiguous* in the sense that they are poorly characterized from a statistical point of view. Such data includes: ambiguous features or attributes; natural-language statements; and rules. The lack of a solid probabilistic foundation for such data has led to the use of a number of heuristic approaches such as fuzzy logic, the Dempster-Shafer theory of evidence, rule-based inference, etc.

The approach is also motivated by the fact that there is no "level playing field" for determining the performance of data fusion algorithms, for comparing one algorithm to another, and so on.

Last but not least, this work is motivated by the fact that *Bayes-optimal multitarget estimation and filtering encounters fundamental conceptual difficulties when the number of targets is unknown*. When one tries to apply the standard statistical thinking just described to the *multitarget case with unknown number of targets*, one quickly discovers that taking things for granted leads to serious troubles that directly bear on practice. First, how do we uniquely specify all of the states that the multitarget system can occupy? Multitarget states must look something like this:

$$\varnothing \quad : \quad \text{no-target state}$$
$$x_1 \quad : \quad \text{the single-target states}$$
$$x_1, x_2 \quad : \quad \text{the two-target states}$$
$$x_1, \ldots, x_t \quad : \quad \text{the } t\text{-target states}$$

However, this specification of states is incomplete. The symbol x, x signifies not that a single target with state x is present twice, but rather that two completely different targets happen to occupy the same kinematic state x. Strictly speaking, therefore, the state of an individual target is not fully specified (in a multitarget context) unless a unique identifier (an "I.D. tag") has been

attached to it, e.g.: (x, τ) where x is the target's kinematic state and τ is its unique identifying tag. Thus the incompletely specified two-target state x, x should be replaced by the completely-specified two-target state (x, τ_1), (x, τ_2). Two-target states of the form (x, τ), (y, τ) with $x \neq y$ must be excluded as non-physical since the same target cannot occupy two different kinematic states simultaneously (see [12], pp. 194-198). Also, note that the two-target state (x_1, τ_1), (x_2, τ_2) and the two-target state (x_2, τ_2), (x_1, τ_1) are not distinct: They represent the same two-target state.

Putting such subtleties aside for the moment, let us proceed immediately to a declaration of victory and assume that we can define a multitarget posterior distribution function on the (thus far vaguely specified) multitarget state space. To keep things simple, assume that all targets are motionless, exist in one dimension, and are completely specified by their locations on the real line as measured in *meters*. Also assume that a single sensor collects a set Z of measurements from the targets, whose number as well as positions are unknown and are to be estimated. One can write a naive posterior distribution function $f(state \mid Z)$ on the multitarget state space, given measurements Z, as follows:

$$f(\varnothing \mid Z) = \text{posterior likelihood of zero targets}$$
$$f(x_1 \mid Z) = \text{posterior likelihood of one target in state } x_1$$
$$f(x_1, x_2 \mid Z) = \text{posterior likelihood of two targets in state } x_1, x_2$$
$$f(x_1, \ldots, x_t \mid Z) = \text{posterior likelihood of } t \text{ targets in state } x_1, \ldots, x_t$$

Since the cumulative likelihood summed over all multitarget states must be 1, it follows that

$$f(0 \mid Z) + f(1 \mid Z) + \ldots + f(t \mid Z) + \ldots + f(M \mid Z) = 1$$

where $f(0 \mid Z) = f(\varnothing \mid Z)$, where $f(t \mid Z) = \int f(x_1, \ldots, x_t \mid Z) \, dx_1 \cdots dx_t$ for $t = 1, \ldots, M$, and where M is the maximum expected number of targets in the scenario. Now, let us naively use the MAP procedure to estimate the complete state of the multitarget system. Then we would write:

$$\hat{x}_1, \ldots, \hat{x}_{\hat{t}} = \underset{t, \, x_1, \ldots, x_t}{argsup} \; f(x_1, \ldots, x_t \mid Z)$$

where $\hat{t}$ is the estimated number of targets and the $\hat{x}_i$ are the estimated positions of the $\hat{t}$ targets. Unfortunately, there is a problem: From the definition of a Riemann integral we know that:

- $f(0 \mid Z) = $ unitless probability
- $f(1 \mid Z) = $ unitless probability, so units of $f(x_1 \mid Z)$ must be *1/meter* since the units of dx_1 are in meters
- $f(2 \mid Z) = $ unitless probability, so units of $f(x_1, x_2 \mid Z)$ must be *1/meter²*
- $f(t \mid Z) = $ unitless probability, so units of $f(x_1, \ldots, x_t \mid Z)$ must be *1/meterᵗ*

6

Consequently the "*argsup*" operation is not defined since the quantities $f(\varnothing \mid Z), f(x_1 \mid Z),...,$ $f(x_1,...,x_t \mid Z),...$ are *incommensurable with respect to units of measurement*. Thus the MAP estimator *cannot even be defined!* One might try to sidestep this problem by using *Riemann-Stieltjes* integrals. That is, let $g(x)$ be an arbitrary density function on (single-target) state space and let $G(x) = \int_0^x g(y) \, dy$ be its corresponding cumulative probability function. Then one could instead define multitarget densities using Riemann-Stieltjes integrals

$$h(t \mid Z) \quad = \quad \int h(x_1,...,x_t \mid Z) \, dG(x_1) \cdots dG(x_t) \quad = \quad \int h(x_1,...,x_t \mid Z) \, g(x_1) \cdots g(x_t) \, dx_1 \cdots$$

where, now, the multitarget distributions

$$h(x_1,...,x_t \mid Z) \quad = \quad \frac{f(x_1,...,x_t \mid Z)}{g(x_1) \cdots g(x_t)}$$

are unitless. Then, the multitarget distribution h will not have the incommensurability-of-units problem just noted and so could be used to define a multitarget MAP estimate. The price, however, is the introduction of an *arbitrary "fudge factor"*--the density g--into the concept of a multitarget posterior distribution h.

If we instead turn to the *posterior expectation* for our salvation, our troubles get even worse. A multitarget posterior expectation, if it exists, must have the general form $\int \langle x_1,...,x_t \rangle f(\langle x_1,...,x_t \rangle \mid Z) \, d\langle x_1,...,x_t \rangle$ where the (as yet to be defined) integral is taken over all (thus far vaguely defined) multitarget states $\langle x_1,...,x_t \rangle$. Such an integral cannot even be defined unless, at minimum, the multitarget state space is a vector space--in particular, unless it has a concept of addition/subtraction. But how does one add the zero-target state $\varnothing$ to a single-target state x? Or a single-target state x to a two-target state x_1,x_2? We could attempt to address this problem by embedding the multitarget state space in a larger, enveloping space which *is* a vector space. In this case, however, there is no guarantee that the posterior expectation would yield values which are actual multitarget states. Rather, it is more likely that it would yield values which are strictly in the enveloping vector space and therefore which would have no physical meaning. *Having been denied the accustomed security of the classical estimators, we are therefore forced to propose new ones and show that they are statistically well-behaved.*

Once such difficulties are exposed, still others are driven into the sunlight. For example, the multitarget posterior densities $f(x_1,...,x_t \mid Z)$ cannot be defined unless one has at hand a multitarget measurement model $f(z_1,...,z_k \mid x_1,...,x_t)$ which tells us the likelihood of seeing measurements $z_1,...,z_k$ given the presence of targets with states $x_1,...,x_t$. Even if a multitarget MAP estimator could be defined, we would still need to know that the multitarget likelihood function is *measurable* in the variables $z_1,...,z_k$ and *continuous* in the variables $x_1,...,x_t$. But measurable with respect to *which topology* on the multitarget measurement space? Continuous with respect to *which metric* on the multitarget state space? The latter question is far from trivial. What is the distance between a single-target state x and a two-target state x_1,x_2? Or the zero-target state $\varnothing$ and the single-target state x_1? What is the distance between the two-target states x_1,x_2 and x_2,x_1 if $x_1 \neq x_2$? (The Euclidean metric gives us $\| (x_1,x_2)-$

$(x_2, x_1) \parallel = \parallel (x_1 - x_2, x_2 - x_1) \parallel \neq 0$. But the state "an F-16 at x_1 flown by Joe and an F-22 at x_2 flown by Ralph" is the same multitarget state as "an F-22 at x_2 flown by Ralph and an F-16 at x_1 flown by Joe".) One can then try to tinker various metrics for multitarget state space, only to get pulled into a a morass of arbitrary, *ad hoc* definitionizing. If these questions are not answered we cannot even *define* a likelihood function: To do so, we need to (1) precisely define the state and measurement spaces, (2) define topologies on the state and measurement spaces, (3) define random variables on the state and measurement spaces using these topologies, and (4) define the multisource, multitarget likelihood function as a conditional distribution in terms of these random variables. Moreover, how can we *systematically construct*, from a knowledge of the Markov state transition models of the individual targets, a Markov sstate-transition for the entire multitarget state--as opposed to just assuming that it comes out of nowhere, *deus ex machina*? Likewise, how can we *systematically construct* the multitarget likelihood function $f(z_1, \ldots, z_k \mid x_1, \ldots, x_t)$ from a knowledge of the characteristics of the individual sensors, without assuming that it, too, comes out of nowhere *deus ex machina*? (Note that in the single-sensor target case we can explicitly construct likelihood functions directly from our knowledge of sensor characteristics. For example, suppose that we know that the sensor is described by the standard Kalman measurement model $z = Cx + w$ where w is random noise with density $f_w(y)$ and C is a matrix. Then the probability measure (mass function) for z is

$$\int_S f_z(y \mid x)\, dy = Pr(z \in S) = Pr(Cx + w \in S) = Pr(w \in S - Cx) = \int_{S-Cx} f_w(y)\, dy = \int_S f_w(y - Cx)\, dy$$

where $S-Cx$ denotes the set of all $s - Cx$ with $s \in S$. Since this is true for all measurable S, the likelihood function is $f_z(y \mid x) = f_w(y - Cx)$ almost everywhere in y. In other words: The likelihood function $f_z(y \mid x)$ is constructed as the Radon-Nikodým derivative $f_z = dp_z/d\lambda$, with respect to Lebesgue measure λ, of probability measure $p_z(S) = Pr(z \in S)$.)

In summary, if we take things for granted and simply declare victory, we are led into fundamental conceptual difficulties that have a direct bearing on engineering practice. What is at issue is neither theoretical hair-splitting nor mere mathematical "bookkeeping." Rather, what is actually at stake is *our ability to do optimal multitarget filtering and estimation at all and, moreover, our ability to even know what "optimal" means in such a context.*

4-A-c. The Need for This Work: USARO Technology Requirements. The proposed work directly addresses the following problem areas described in the 1995 USARO Broad Area Announcement:

● "...determining information-theoretic bounds on the performance of any algorithm, given parameters of the sensor and the scene." (p. 19)

● "Processes such as target detection, target recognition and classification, and tracking often require fusion of information from potentially diverse sources. The increasing volume and variety of battlefield data makes the "higher" levels of information fusion-- situation and threat assessment--increasingly crucial. Decision-making at these higher

8

fusion levels often involves forms of ambiguity more extreme than those addressed by conventional statistical analysis; imprecision, vagueness, indiscernability, etc. Therefore, it is essential that information from sources such as images, signals, voice messages, geographical information, natural language text, and prior knowledge/rule bases be presented in a unified framework. Mathematical methods for measuring and representing information content in diverse and ambiguous data are fundamental. Recent approaches (Dempster-Shafer, fuzzy logic, rule-based, rough sets, statistical capacities, etc.) offer a way of dealing with highly ambiguous information. A systematic, tractable framework is needed that will allow diverse input data streams to be transformed into a unified information fusion space. Such measures should enable prediction of the level of system performance achievable based upon the information content of sources available, knowledge gained from previous experience, tasks o be performed, and constraints in the context of the task to be performed." (p. 19)

● "In higher levels of information fusion such as situation and threat assessment and mission management, the primary objective is to simulate a human expert (e.g. rule-based system, adaptive system, neural networks, etc.). Issues which require further research include: more systematic approaches to situation assessment which permit effecive performance analysis, prediction, and evaluation; and assessment of measures of stability and performance. A common methodology should be developed that would support the optimal, adaptive management of information collection resources." (p. 20)

● "The fundamental concept of a system that can process artificially sensed information, make optimal decisions based on this information and well-defined objecives and translate these decisions into actions is a guiding and unifying theme for basic research in all major aspects of this area [Foundations of Intelligent Control Systems]." (p. 49)

In addition, the proposed directly and explicitly addresses the following requirements set forth in the report of the USARO Myrtle Beach Electronics Strategy Planning Workshop, Jan. 9-12 1995:

Thrust IA-7 Mathematical Representations and Measures to Unify Decision-Making Based on Diverse Information (Extremely High Priority). "Processes, such as target detection, target recognition and classification, and tracking often require fusion of information from potentially diverse sources. The increasing volume and variety of battlefield data makes the higher levels of information fusion--situation and threat assessment, mission management--increasingly crucial. Decisionmaking at these higher fusion levels often involves forms of ambiguity more extreme than those addressed by conventional statistical analysis: imprecision, vagueness, indiscernability, partial contingency, etc. Therefore, it is essential that information from sources such as images, signals, voice messages, geographical information, natural language text, and prior knowledge/rule bases be presented in a unified framework. Mathematical methods for measuring and representing information content in diverse and ambiguous data are fundamental. Classical statistical decision theory provides methods for dealing with

uncertain information when the underlying probabilities are known or can be treated subjectively. Recent approaches (Dempster-Shafer, fuzzy logic, rule-based, rough sets, statistical capacities, etc.) offer a way of dealing with highly ambiguous information. These methods utilize a variety of information-quality measures: possibility, belief, entropy, etc. A systematic, tractable framework is needed that will allow diverse input data streams to be transformed into a unified information fusion space. This framework should provide systematic, tractable measures of information quality. Such measures should enable prediction of the level of system performance achievable (see Thrust IA-4) based upon the information content of sources available, knowledge gained from previous experience, tasks to be performed, and constraints in the context of the task to be performed." (pp. 40-41)

Thrust IA-8 Multi-Criterion, Multi-Expert, and Multi-Source Decision Analysis (Very High Priority). "In conventional approaches to optimization, a key assumption is that the performance of a system can be assessed by a single criterion, e.g. cost. In many real-world situations this is not the case. Furthermore, performance assessments, decisions and/or estimates may be provided by a number of experts or fusion sources, each employing different evaluation criteria and using possibly overlapping data sources. A common examples [sic] is "track-to-track" fusion, in which existing and possibly correlated or conflicting estimates/decisions must be fused into a valid composite picture. Available methods are hard to apply and are lacking in computational efficiency. Techniques are needed which are capable of representing preferences, expert credibilities, weights of criteria importance, and data dependencies in qualitative terms that lead to an aggregated choice of alternatives which are preferable or admissible but not necessarily optimal." (p. 41)

Thrust IA-9 Methodologies for Situation Analysis (High Priority). "In higher levels of information fusion such as situation and threat assessment and mission management, the primary objective is a simulation of a human expert (e.g. rule-based system, adaptive system, neural networks, etc.) which approximates the expertise of the human. Issues which require further research include: more systematic approaches to situation assessment which permit effective performance analysis, prediction, and evaluation; and assessment of measures of stability and performance. A common methodology should be developed that would support the optimal, [sic] of information resources." (p. 42)

4-B. SUMMARY OF MOST IMPORTANT RESULTS

4-B-a. Background and Objectives. In this section we will summarize the major findings of the work completed thus far under contract DAAH04-94-C-0011. The basic problems of data/information fusion are summarized in **[1, 17, 60]**. During the current project Lockheed Martin constructed a rigorous, fully probabilistic scientific foundation for the following aspects of information fusion:

(1)   Multisource integration based on parametric estimation and Markov techniques **[36, 38,**

40, 42, 43, 44]

(2) Prior information regarding the numbers, identities, and geokinematics of targets [36]

(3) Sensor management based on information theory and nonlinear control theory [32]

(4) Performance evaluation using information theory and nonparametric estimation [39, 41]

(5) Expert-systems theory: fuzzy logic, evidential theory, rule-based inference [29, 30, 35]

The existence of this unification approach implies the existence of algorithms which fully unify, in *a single statistical process*, the following functions of information fusion:

- Target detection
- Target identification
- Target tracking and localization
- Prior information with respect to detection, classification, and tracking
- Ambiguous *evidence* and precise *data*
- Evidential rules of combination
- Sensor management, including optimal selection of sensor-control parameters

Not all sensors can be *directly* subsumed within this unification, however. We assume the following sensor types: (1) point-source, (2) range-profile, (3) line-of-bearing, (4) human observers reporting in natural language, (5) rulebases, (6) imaging sensors whose target-images are point "firefly" sources, (7) imaging sensors whose target-images consist of relatively small clusters of point energy reflectors (also sometimes called "extended targets").

Random set theory was developed by Kendall [25] and Mathéron [45] in the context of stochastic geometry. Since the late 1970s, several researchers have investigated the connections between random sets and fuzzy logic [9, 11, 51], the Dempster-Shafer evidential theory [8, 18, 31, 37, 49, 57], rule-based inference [33, 34], general expert systems theory [13, 15, 26, 53], and data fusion [10, 47, 48]. The work performed under the current contract has built upon and greatly extended this existing body of research, resulting in a systematic and general mathematical apparatus for solving information fusion problems.

The basic approach is as follows. A suite of known sensors transmits to a *central data fusion site* the observations they collect regarding a group of targets whose numbers, positions, velocities, identities, threat states, etc. are unknown. Finite-set statistics arises when we *mathematically* reformulate the multisensor, multisource problem as a single-sensor, single-source problem:

- A group of sensors $\{\odot_1,\ldots,\odot_s\}$ reporting to a central data fusion site is modeled as a single "global" sensor $\odot*$

- An unknown number of targets $\{\otimes_1,\ldots,\otimes_t\}$ is modeled as a single "global" target $\otimes*$

- A group of reports $\xi_1,\ldots,\xi_r$, collected when the sensors interrogate the targets at a given instant, is modeled as a single "global" report $Z = \{\xi_1,\ldots,\xi_r\}$

Simplifying somewhat, each individual target state will have the form

$$\zeta = (\mathbf{x}, v, j)$$

where $\mathbf{x}$ is a continuous state vector (e.g. geokinematics), v is a discrete state variable (e.g. target class) and j is a (generally unknown and unknowable) unique identifying tag associated with each target. The complete state of a multitarget system is, therefore, represented by a "set parameter" of the form

$$X = \{\zeta_1, \ldots, \zeta_v\}$$

which is, in general, a specific value of a *randomly varying* parameter-set Γ.

Likewise, each individual observation will have the form

$$\xi = (\mathbf{z}, u, i)$$

where $\mathbf{z}$ is a continuous variable (geokinematics, signal intensity, etc.) in $\mathbb{R}^n$, where u is a discrete variable (e.g. possible target attributes) drawn from a finite universe U of possibilities, and where i is a "sensor tag" which identifies the sensor which supplied the measurement. If the total observation-set (the global observation)

$$Z = \{\xi_1, \ldots, \xi_k\}$$

collected by the global sensor is treated as a single entity, then it is a specific value of a *randomly varying* finite observation-set Σ.

4-B-b. Random Set Formulation of Data Fusion Problems. The global report, which varies randomly in regard to kinematics, identities, and numbers of elements, is a *finite random set*, say Σ. Given this, it becomes possible in principle to *reformulate the multisensor, multitarget data fusion problem as a single-sensor, single-target problem*. The statistics of the finite random set Σ are determined not by a probability measure but rather by the "belief measure" $\beta_\Sigma(S \mid X)$ $= p(\Sigma \subseteq S)$ which belongs to a family of belief measures parametrized by a parameter X which consists of a *finite subset* of (conventional) parameters.

In more detail, from Mathéron's random set theory we know that the class of finite subsets of measurement space has a topology, the so-called *hit-or-miss topology* [45,46]. If O is any Borel subset of this topology then the statistics of Σ are characterized by the associated probability measure

$$p_\Sigma(O) = p(\Sigma \in O)$$

The Choquet-Mathéron capacity theorem (see [45], pg. 30) tells us, among other things, that we need only consider Borel sets of the specific form

12

$$O_{S^c}^c$$

i.e., the class whose elements are all closed subsets C of measurement space such that $C \cap S^c \neq \varnothing$ (i.e., $C \subseteq S$) where S is some closed subset of measurement space. In this case

$$p_\Sigma(O_{S^c}^c) \quad = \quad p(\Sigma \subseteq S) \quad = \quad \beta_\Sigma(S)$$

The set function β_Σ is known as the *belief measure* of the randomly varying finite subset Σ, or alternatively as an "infinitely monotone capacity." Whereas a probability measure q is "additive" (i.e., $q(S \cup T) = q(S) + q(T)$ if $S \cap T = \varnothing$) a·belief measure is, in general, nonadditive: $\beta(S \cup T) > \beta(S) + \beta(T)$ if $S \cap T = \varnothing$. Thus we can use the *nonadditive* measure β_Σ defined on subsets of *ordinary observation space* instead of the *additive* measure p_Σ, which is defined on subsets of the class of finite sets of observation space.

Despite the fact that the belief measure β_Σ is nonadditive, it plays the same role in multisensor, multitarget statistics that ordinary probability measures play in single-sensor, single-target statistics. The reason is that from the belief measure β_Σ we can construct a "global density function" $f_\Sigma(Z)$ which describes the *comprehensive statistical behavior of the entire sensor suite*. Just as the density of a random vector can be derived from its cumulative probability function through iterated differentiation, so the *global density* of a finite random set can be derived from its belief measure via iteration of a generalized form of the familiar Radon-Nikodým derivative from Lebesgue measure theory.

In what follows, $\mathfrak{R} \triangleq \mathbb{R}^n \times U \times \{1,...,s\}$ denotes a "hybrid space" of continuous and discrete observation variables, where U is some finite set. That is, a typical element $\xi = (z,u,j) \in \mathfrak{R}$ is a triple which denotes a continuous (e.g. geopositional) observation $z \in \mathbb{R}^n$, a discrete observation $u \in U$ (e.g. a target identity, target class, or other discrete identity-type attribute), and $1 \leq j \leq s$ is an integer "sensor tag" which identifies the sensor which collected the continuous-discrete observation (z,u). Let $Z = \{\xi_1,...,\xi_k\}$ be a finite subset of $\mathfrak{R}$ with $\xi_1,...,\xi_k$ being distinct.

4-B-c. An Integral and Differential Calculus for Data Fusion.

The theoretical core of the random set approach is a generalization of integral and differential calculus to certain types of set functions (e.g. belief measures). The *set integral* and its inverse operation the *set derivative* play somewhat the same role in multisensor, multitarget problems that the conventional integral and derivative play in single-sensor, single-target problems.

If an *additive* measure $p_Z(S) = p(Z \in S)$ of a random vector Z is absolutely continuous with respect to Lebesgue measure λ then by the Radon-Nikodým theorem one can, in principle, determine the *density function*

$$f_Z = \frac{dp_Z}{d\lambda}$$

that corresponds to it. Conversely, the measure p_Z can be recovered from the density through application of the Lebesgue integral

$$\int_S f_Z(z) \, d\lambda(z) = p_Z(S)$$

We have shown how to define an integral and differential calculus of functions of a set variable which obeys similar properties. Given a vector-valued function $f(Z)$ of a *finite-set* variable Z we have shown how to define a "set integral" of the form $\int_S f(Z) \, \delta Z$. Conversely, given a vector-valued function $\Phi(S)$ of a *closed set* variable S, we have shown how to *constructively* define a "set derivative" of the form $\delta \Phi / \delta Z$. Under certain assumptions, these operations turn out to be inverse to each other:

$$\int_S \frac{\delta \Phi}{\delta Z}(\varnothing) \, \delta Z = \Phi(S)$$

$$\left[\frac{\delta}{\delta Z} \int_S f(Z) \, \delta Z \right]_{S=\varnothing} = f(Z)$$

In more detail, the global density $f_\Sigma(Z)$ is derived from the belief measure $\beta_\Sigma(S \mid X)$ by an iterated differentiation:

$$f_\Sigma(Z) = \frac{\delta^k \beta_\Sigma}{\delta \xi_1 \cdots \delta \xi_k}(\varnothing)$$

This differentiation, called "generalized Radon-Nikodým differentiation," is defined by

$$\frac{\delta \beta_\Sigma}{\delta \xi}(T) = \lim_{E \to \{z\}} \frac{\beta_\Sigma(T \cup (E \times \{u\} \times \{j\})) - \beta_\Sigma(T)}{\lambda(E)}$$

for any $\xi = (z, u, j) \in \Re$, where E is a closed ball in $\mathbb{R}^n$ centered at z which converges to the point-set $\{z\}$, and where $\lambda(E)$ denotes the Lebesgue measure (hypervolume) of E in $\mathbb{R}^n$. (This is a somewhat simplified definition, assuming $\xi \notin T$. Also, the limit as E approaches the singleton set $\{z\}$ requires careful treatment based on the constructive definition of the Radon-Nikodým derivative [56, 61], as described in section 4.2.3 of Chapter 4 of [12]). The belief measure can be recovered from the global density via integration,

$$\beta_\Sigma(S) = \int_S f_\Sigma(Z) \, \delta Z$$

14

which, setting $S = \Re$, yields the normality condition $\int f_\Sigma(Z) \, \delta Z = 1$. Here, the integral is a so-called *set integral*, which is defined as follows. Let $\Phi(Z)$ be a function whose arguments are finite sets Z. Then

$$\int_S \Phi(Z) \, \delta Z = \sum_{k=0}^\infty \frac{1}{k!} \int_{S^k} \Phi(\{\xi_1, ..., \xi_k\}) \, d\lambda(\xi_1) \cdots d\lambda(\xi_k)$$

where the integrals on the right-hand side of the equation are "hybrid integrals" and where $S^k = S \times ... \times S$ denotes the Cartesian product of S taken with itself k times. The "hybrid integral" (i.e., the integral defined in terms of the product measure of Lebesgue measure on $\mathbb{R}^n$ and the counting measure on U) of a function $\phi(\xi)$ whose arguments are $\xi = (z, u) \in \Re$ is defined by:

$$\int_S \phi(\xi) \, d\lambda(\xi) = \sum_{u \in U} \int_{S(u)} \phi((z, u)) \, d\lambda(z)$$

for any $S \subseteq \Re$, where $S(u)$ denotes the subset of all $z \in \mathbb{R}^n$ such that $(z, u) \in S$.

4-B-d. The Global Density of a Sensor Suite. Given certain absolute continuity assumptions which need not concern us here, if β_Σ is the belief measure of a random observation set Σ then the quantity

$$f_\Sigma(Z) = \frac{\delta \beta_\Sigma}{\delta Z}(\varnothing)$$

is the *density function* of the random finite subset Σ. (This is essentially equivalent to what, in point process theory, are called "Janossy densities.") The quantity $f_\Sigma(Z)$ is the likelihood (i.e., probability density) that the event $\Sigma = Z$ will occur. On the other hand, $f_\Sigma(Z)$ also has a completely Bayesian interpretation. Suppose that $f_\Sigma(Z \mid X)$ is a global density with a *set parameter* $X = \{\zeta_1, ..., \zeta_t\}$, where $\zeta_1, ..., \zeta_t$ are the unknown (discrete and continuous) parameters of the targets, and t is the unknown number of the targets. Then $f_\Sigma(Z \mid X)$ is the *total probability density of association* between the measurements in Z and the parameters in X.

The global density of a sensor suite differs from conventional densities in that it *encapsulates the comprehensive statistical behavior of the entire sensor suite into a single mathematical object*, and not just sensor noise statistics. Most generally speaking, a global density has the form $f_\Sigma(Z \mid X, Y)$ and includes the following information:

- the observation-set $Z = \{\xi_1, ..., \xi_k\}$
- the set $X = \{\zeta_1, ..., \zeta_t\}$ of unknown parameters
- the states $Y = \{\eta_1, ..., \eta_s\}$ of the sensors (orientations, modes, etc.)
- the sensor-noise distributions of the individual sensors
- the probabilities of detection and false alarm for the individual sensors

15

- clutter models
- detection profiles for the individual sensors (as functions of range, aspect, etc.)

Global density functions can be computed explicitly from a knowledge of the sensors that belong to the sensor suite. For example, suppose that we are given a single sensor with (conventional) sensor-noise density $f(\xi \mid \zeta)$ with no false alarms and constant probability of detection p_D, and that observations are independent. Then the global density which specifies the *multitarget measurement model* for the sensor can be shown to be

$$f_\Sigma(\{\xi_1,...,\xi_k\} \mid \{\zeta_1,...,\zeta_t\}) = p_D^k(1-p_D)^{t-k} \sum_{1 \le i_1 \ne ... \ne i_k \le t} f(\xi_1 \mid \zeta_{i_1}) \cdots f(\xi_k \mid \zeta_{i_k})$$

where the summation is taken over all distinct $i_1,...,i_k$ such that $1 \le i_1,...,i_k \le t$. (Here, $\xi_1,...,\xi_k$ are assumed distinct, as are $\zeta_1,...,\zeta_t$.)

Global densities satisfy the properties that one would expect. For example, if $\mathbf{T}(Z)$ is a measurable vector-valued transformation of a finite-set variable then the expectation $E[\mathbf{T}(\Sigma)]$ of the random vector $\mathbf{T}(\Sigma)$ is

$$E[\mathbf{T}(\Sigma)] = \int \mathbf{T}(Z) f_\Sigma(Z) \, \delta Z$$

where the integral is a set integral. Global densities also enjoy certain *unexpected* properties, the most interesting of which is the fact that they are *continuous analogs of the Möbius transform of Dempster-Shafer theory*. That is, if $Z = \{z_1,...,z_k\}$ with $|Z| = k$ then it can be shown that

$$f_\Sigma(Z) = \lim_{i \to \infty} \frac{\sum_{Y \subseteq Z} (-1)^{|Z-Y|} \beta_{\Sigma,i}(Y)}{\lambda(E_{z_1;i}) \cdots \lambda(E_{z_k;i})}$$

where $\beta_{\Sigma;i}$ is defined by

$$\beta_{\Sigma,i}(\{z_1,...,z_k\}) = \beta_\Sigma(E_{z_1;i} \cup ... \cup E_{z_k;i})$$

where $E_{z;i}$ denotes a closed ball of radius $1/i$ centered at $z \in \mathbb{R}^n$.

4-B-e. A Simple Illustration. Assume that two identical targets located on the real line are observed by a single Gaussian sensor with Gaussian density

$$f(a \mid x) = N_{\sigma^2}(a-x)$$

with variance σ^2 and associated probability measure $p_f(S \mid x)$. Assume also that (1) reports are independent, (2) the probability of false alarm for the sensor is zero, (3) the probability of detection q is not necessarily unity, and (4) that q is constant over the region of interest. The

16

observations corresponding to the first target are the specific realizations of a Gaussian random number A_1 and the observations corresponding to the second target are modeled by another Gaussian random number A_2. If $q = 1$ then the observation generated by the first target is a randomly-varying single-element set of the form $\Sigma_1 = \{A_1\}$ and the observation generated by the second target is $\Sigma_2 = \{A_2\}$. What the sensor sees at any given instant is just an unordered two-element set $Z = \{a_1, a_2\}$ of reports generated by the randomly varying two-element set $\Sigma = \Sigma_1 \cup \Sigma_2 = \{A_1, A_2\}$. If $q < 1$, however, then the possible realizations for Σ are $\Sigma = Z$ where 1) $Z = \{a_1, a_2\}$ for any a_1, a_2 in $\mathbb{R}$ with $a_1 \neq a_2$ 2) $Z = \{a\}$ for any a in $\mathbb{R}$; or 3) $Z = \varnothing$.

The statistics of this random set are described by the belief measure

$$\beta_\Sigma(S \mid X) \;=\; p(\Sigma \subseteq S) \;=\; p(\Sigma_1 \subseteq S, \; \Sigma_2 \subseteq S)$$

where X is a set parameter which can take the form $X = \varnothing$ (no tracks), $X = \{x\}$ (one track), or $X = \{x_1, x_2\}$ with $x_1 \neq x_2$ (two tracks). In fact,

$$\beta_\Sigma(S \mid \{x\}) \;=\; 1-q + q\, p_f(S \mid x)$$

$$\beta_\Sigma(S \mid \{x_1, x_2\}) \;=\; [1-q + q\, p_f(S \mid x_1)]\,[1-q + q\, p_f(S \mid x_2)]$$

From the belief measure it is possible, using the set derivative, to compute the global density. In the one-track case $X = \{x\}$ it is

$$f_\Sigma(\varnothing \mid X) \;=\; 1-q, \qquad f(\{z\} \mid X) \;=\; q\, f(z \mid x)$$

and in the two-track case $X = \{x_1, x_2\}$ by

$$f_\Sigma(\varnothing \mid X) \;=\; (1-q)^2$$

$$f(\{z\} \mid X) \;=\; q(1-q)\, f(z \mid x_1) + q(1-q)\, f(z \mid x_2)$$

$$f(\{z_1, z_2\} \mid X) \;=\; q^2 f(z_1 \mid x_1)\, f(z_2 \mid x_2) + q^2 f(z_2 \mid x_1)\, f(z_2 \mid x_2)$$

In all other cases, $f_\Sigma(Z \mid X) = 0$. The one-track case is just the conventional single-sensor, single-target situation with the probability of detection taken into explicit account ([47] p. 15, [48]). The two-track case is more interesting. Given any real numbers $a_1 < a_2$ the set $\{a_1, a_2\} = \{a_2, a_1\}$ does not depend on the order of a_1, a_2 and therefore can be identified with the point (a_1, a_2) on the half-plane. Likewise, when $a_1 = a_2 = a$ then the set $\{a\}$ can be identified with the diagonal boundary line $x = y$ of the half-plane. In other words, all of the possible two-observation realizations $Z = \{a_1, a_2\}$ of the random set Σ can be identified with points on the half-plane. The global density $f_\Sigma(\{a_1, a_2\} \mid X)$ gives the probability density that $\{a_1, a_2\}$ occurs as a value of the random set $\{A_1, A_2\}$. Its graph is just a surface over the half-plane which has a peak near (but usually not the same as) the point (a_1, a_2). The reader

may also easily verify that the belief measure $\beta_\Sigma(S \mid X)$ can be recovered from the global density via the set integral:

$$\beta_\Sigma(S \mid X) \;=\; \int_S f_\Sigma(Z \mid X)\,\delta X$$

$$=\; f_\Sigma(\varnothing \mid X) \;+\; \int f_\Sigma(\{z\} \mid X)\,dz \;+\; \tfrac{1}{2}\int_{S \times S} f_\Sigma(\{z_1, z_2\} \mid X)\,dz_1\,dz_2$$

for any closed $S \subseteq \mathbb{R}$ and any finite $X \subseteq \mathbb{R}$.

4-B-f. The Parallelism Between Point- and Finite-Set Statistics. Because of the existence of the set derivative and the set integral, one can compile the following list of *direct mathematical parallels* between the single-sensor, single-target random-vector world (conventional statistics) and the multisensor, multitarget finite random set world ("finite-set statistics"). These parallels can be expressed as "translation dictionary":

Random Vector, $\mathbf{Z}$	Finite Random Set, Σ
sensor, $\odot$	global sensor, $\odot*$
target, $\otimes$	global target, $\otimes*$
report, $\mathbf{z}$	global set report, Z
vector parameter, $\mathbf{x}$	set parameter, X
differentiation, $dp_Z/d\mathbf{z}$	set differentiation, $\delta\beta_\Sigma/\delta Z$
integration, $\int_S f_Z(\mathbf{z} \mid \mathbf{x})\,d\lambda(\mathbf{z})$	set integration, $\int f_\Sigma(Z \mid X)\,\delta Z$
probability measure, $p_Z(S \mid \mathbf{x})$	belief measure, $\beta_\Sigma(S \mid X)$
density, $f_Z(\mathbf{z} \mid \mathbf{x})$	global density, $f_\Sigma(Z \mid X)$
prior density, $\pi(\mathbf{x})$	global prior density, $\pi(X)$
estimators	global estimators
information (entropy) metrics	global information metrics
Markov transition densities	global Markov transition densities
nonlinear filtering	global nonlinear filtering
control theory	global control theory

This parallelism between the information fusion and single-sensor, single-target tracking worlds means that general statistical methodologies can, with a bit of work, be directly "translated" from the random-vector case to the random-set case. That is, any theorem or mathematical algorithm in conventional statistics can be thought of as a "sentence" in a language whose "words" and "grammar" consist of the basic concepts in the left-hand column above. The above two columns can be thought of as a "dictionary" which establishes a direct correspondence between the words and grammar in the random-vector language and the cognate words and grammar of the random-set language. Consequently, any "sentence"--any theorem or mathematical algorithm--phrased in the random-vector language can, in principle, be directly "translated" into a corresponding sentence (and thus corresponding theorem or algorithm) in the

18

random-set language. This process of direct translation can be encapsulated as a general principle:

> **The "Almost-Parallel Worlds Principle" (APWOP)**: *Nearly every theorem, method, or algorithm for single-sensor, single-target tracking implies an analogous theorem, method, or algorithm in multi-sensor, multi-target data fusion*

We say "almost-parallel" because, as with any translation process, the correspondence between dictionaries is not precisely one-to-one. For example, there apparently is no natural way to add and subtract finite sets as one does vectors. Nevertheless, the parallelism is complete enough that, provided one exercises some care, a hundred years of accumulated knowledge in conventional (i.e., single-sensor, single-target) statistics can be *directly* brought to bear on multisensor, multitarget information fusion problems. In particular, we offer the following specific examples of the potential utility and power of the "almost-parallel worlds principle."

a. Data Fusion Information Metrics. Suppose that we wish to attack the problem of *performance evaluation* of information fusion algorithms in a scientifically defensible manner. In the proposal which led to this contract, we defined a "global" version of the conventional Kullback-Leibler discrimination (or cross-entropy) metric which is applicable to performance evaluation problems in multisensor, multitarget data fusion:

$$I(f_\Sigma; f_\Gamma) \quad = \quad \int f_\Sigma(X) \, \ln\!\left(\frac{f_\Sigma(X)}{f_\Gamma(X)}\right) \, \delta X$$

where the integral is a *set integral*. This information metric is the basis for a systematic, information-theory based approach to multisource, multitarget data fusion system evaluation. Using such metrics, one can measure the overall information produced by a data fusion system, or the specific "components" of information attributable to such special functions as target I.D., target localization, or target detection. In addition, a well-known challenge in performance measurement is the fact that one end-user's "information" is another end-user's "confusion." Accordingly, information-theory based approaches to performance evaluation will not prove adequate unless ambiguous or even subjective *user-specified* definitions of information can be modeled and computed. Under the current contract, it was shown that this is indeed possible. (See section 5.3 of Chapter 5 of **[12]**, and section 8.1 of Chapter 8 of **[12]** for more details.)

b. Multisensor, Multitarget Decision Theory. The basic elements of decision theory can be directly generalized to the multisource, multitarget case. For example, the *Receiver Operating Characteristic (ROC) curve* of an entire multisensor, multitarget system can be defined as the parametrized curve $\tau \to (p_{FA}(\tau), p_D(\tau))$ where

$$p_{FA}(\tau) = \int_{L(Z_1,...,Z_m) > \tau} f_{\Sigma_1,...,\Sigma_m | H_1}(Z_1,...,Z_m | H_1)\, \delta Z_1 \cdots \delta Z_m$$

$$p_D(\tau) = 1 - \int_{L(Z_1,...,Z_m) < \tau} f_{\Sigma_1,...,\Sigma_m | H_0}(Z_1,...,Z_m | H_0)\, \delta Z_1 \cdots \delta Z_m$$

and where

$$L(Z_1,...,Z_m) = \frac{f_{\Sigma_1,...,\Sigma_m | H_0}(Z_1,...,Z_m | H_0)}{f_{\Sigma_1,...,\Sigma_m | H_1}(Z_1,...,Z_m | H_1)}$$

is the "global" likelihood ratio for the problem of deciding between hypotheses H_0 and H_1 given global observations $Z_1,...,Z_m$. (See [28] for more details.)

c. Multisensor, Multitarget Estimation. One can define multisensor, multitarget analogs of conventional statistical estimators, as well as the concept of a Bayes-optimal multisensor,multitarget estimator. Each such estimator is actually a different possible *optimal multisensor, multitarget Level 1 data fusion algorithm.*

Specifically, a *global estimator* of the set parameter X of a global density $f_\Sigma(Z \mid X)$ is a finite-set-valued function $\hat{X} = J(Z_1,...,Z_m)$ of the global measurements $Z_1,...,Z_m$. One can define a multisensor, multitarget version of the maximum likelihood estimator (MLE) by

$$J_{GMLE}(Z_1,...,Z_m) = arg\ sup_X L(X \mid Z_1,...,Z_m)$$

where $L(X \mid Z_1,...,Z_m) = f_\Sigma(Z_1 \mid X) \cdots f_\Sigma(Z_m \mid X)$ is the "global" likelihood function. That is, one determines that value of t and those values of $\zeta_1,...,\zeta_t$ which maximize the value of the likelihood function $L(\{\zeta_1,...,\zeta_t\} \mid Z_1,...,Z_m)$.

Notice that in determining X, one is estimating not only the geokinematics and identities of targets, but also their number as well. Thus detection, localization, and identification are unified into a single statistical operation. This operation is a *direct* multisource, multitarget estimation technique in the sense that the unknown states of the unknown number of targets are determined directy from data, without first attempting to compute an optimal report-to-track assignment. The definition of a multisensor, multitarget analog of the MAP estimator is less straightforward but also possible. See section 5.2 of Chapter 5 of [12] for more details.

d. Cramér-Rao Inequalities. One can show, under assumptions analogous to those used in the proof of the conventional Cramér-Rao inequality, that the following holds for a *vector-valued* multisensor, multitarget estimator **J**:

$$\langle v, C_{J,X}(v)\rangle \cdot \langle w, L_{X,x}(w)\rangle \geq \langle v, \frac{\partial}{\partial_x w} E_X[X]\rangle$$

for all v, w where $X = J(\Sigma_1, \ldots, \Sigma_m)$ and where $L_{X,x}$ is defined by

$$\langle v, L_{X,x}(w)\rangle = E_X\left[\left(\frac{\partial \ln f}{\partial_x v}\right)\left(\frac{\partial \ln f}{\partial_x w}\right)\right]$$

for all v, w, where

$$f = f_{\Sigma_1, \ldots, \Sigma_m | r}$$

and where the directional derivative $\partial f / \partial_x v$ of the function $f(X)$ of a finite-set variable X, if it exists, is defined by

$$\frac{\partial f}{\partial_x v}(X) = \lim_{\varepsilon \to 0} \frac{f((X - \{x\}) \cup \{x + \varepsilon v\}) - f(X)}{\varepsilon} \quad (\text{if } x \in X)$$

$$\frac{\partial f}{\partial_x v}(X) = 0 \quad (\text{if } x \notin X)$$

The existence of this multisource, multitarget Cramér-Rao inequality opens the possibility of determining best-possible-theoretical performance bounds for specific multisource, multitarget data fusion problems. See section 5.3 of Chapter 5 of [12] for more details.

e. Multisensor, Multitarget Nonlinear Filtering. Moving targets can be accounted for through suitable generalization of standard Bayes-Markov filtering techniques. Let $Z_1, \ldots, Z_\alpha$ be a time-sequence of global observation-sets collected by the global sensor and abbreviate the list $Z_1, \ldots, Z_\alpha$ as $Z^{(\alpha)}$. Let $f(Z_\alpha | X_\alpha)$ be the global density which describes the behavior of the global sensor. Under Markov assumptions, the dynamic time-evolution of *the entire multitarget system* between measurements is described by a *global* Markov transition density $f_{\alpha+1 | \alpha}(X_{\alpha+1} | X_\alpha)$. Note that, in general, the finite parameter-sets $X_{\alpha+1}$ and X_α may have differing numbers of elements. This fact allows for our global motion model to model variation in the number of targets due to appearance and disappearance of targets.

Lockheed Martin has shown that it is possible to contruct global transition densities $f_{\alpha+1 | \alpha}(X_{\alpha+1} | X_\alpha)$ from the motion models $f_{\alpha+1 | \alpha}(\zeta_{\alpha+1} | \zeta_\alpha)$ of individual targets in such a way that the global-target state transition is completely consistent with the state-transitions of the individual targets. See section 6.4.3 of Chapter 6 of [12]. As a special case, assume that the number of targets always remains constant over time, then a typical global transition density will have the form

$$f_{\alpha+1|\alpha}(\{\xi_1,...,\xi_t\} \mid \{\zeta_1,...,\zeta_t\}) = \sum_{\sigma} f_{\alpha+1|\alpha}(\xi_1 \mid \zeta_{\sigma 1}) \cdots f_{\alpha+1|\alpha}(\xi_t \mid \zeta_{\sigma t})$$

for any $t = 0,1,2....$ where $f_{\alpha+1|\alpha}$ is a conventional Markov transition density and σ is a permutation on the numbers $1,...,t$. The global transition density formulation is general enough in form that it can accommodate different motion models for different targets. The reason is that, if $\xi_{\alpha+1} = (\mathbf{x}_{\alpha+1}, u)$ and $\zeta_\alpha = (\mathbf{x}_\alpha, u)$ then $f_{\alpha+1|\alpha}(\xi_{\alpha+1} \mid \zeta_\alpha) = f_{\alpha+1|\alpha;u}(\mathbf{x}_{\alpha+1} \mid \mathbf{x}_\alpha)$ where $f_{\alpha+1|\alpha;u}$ denotes a Markov transition density associated with the target or target type identified by u.)

It follows that the concept of Markov time-prediction can be directly generalized to global densities:

$$f_{\alpha+1|\alpha}(X_{\alpha+1} \mid Z^{(\alpha)}, U^\alpha) = \int f_{\alpha+1|\alpha}(X_{\alpha+1} \mid X_\alpha, \mathbf{u}_\alpha) f_{\alpha|\alpha}(X_\alpha \mid Z^{(\alpha)}, U^{\alpha-1}) \, \delta X_\alpha$$

where the integral is a set integral and where U^α and $\mathbf{u}_\alpha$ are control inputs. Likewise, the discrete-time Bayesian nonlinear filtering equation can be directly generalized:

$$f_{\alpha+1|\alpha+1}(X_{\alpha+1} \mid Z^{(\alpha+1)}, U^\alpha) = \frac{f(Z_{\alpha+1} \mid X_{\alpha+1}) f_{\alpha+1|\alpha}(X_{\alpha+1} \mid Z^{(\alpha)}, U^\alpha)}{\int f(Z_{\alpha+1} \mid Y_{\alpha+1}) f_{\alpha+1|\alpha}(Y_{\alpha+1} \mid Z^{(\alpha)}, U^\alpha) \, \delta Y_{\alpha+1}}$$

Estimates $\hat{X}_\alpha$ of the time-evolving state X_α can be computed using using estimation approaches (e.g., MAP estimation: determining which value of the set parameter X_α maximizes the global posterior (in a specific sense not discussed here) at any given time-instant). For more details, see section 6.4 of Chapter 6 of [12].

f. Sensor Management: Multisensor Nonlinear Control Theory. Recall that *sensor management* is the problem of controlling the re-directable and/or multimode sensors in a sensor suite in order to resolve ambiguities in our knowledge about multiple targets. The parallelism between point-variate and finite-set-variate statistics suggests that one way of attacking this problem is by first examining the solution methodology for the *single-sensor, single-target* case: E.g., a missile-tracking camera as it attempts to follow a missile. The camera must adjust its azimuth, elevation, and focal length in such a way as to anticipate the location of the missile at the time the next image of the missile is to be recorded. This is a standard problem in *optimal control theory*. Such problems are solved by defining a *controlled vector*, associated with the camera, and a *reference vector*, associated with the target, and attempting to keep the distance between these two vectors as small as possible.

An approach to the multisensor, multitarget sensor management problem becomes evident if we use the "almost-parallel worlds principle" to reformulate such problems as a single-sensor, single-target problem. In this case the "global" sensor follows a "global" target (even though some of whose individual targets may not even be detected yet). The motion of the multitarget

system is modeled using a global Markov transition density. The only undetermined aspect of the problem is how to define analogs of the controlled and reference vectors. This is done by determining the Kullback-Leibler information distance between two suitable global densities.

In more detail, suppose that the global sensor is "following" the "trajectory" of the global target, in some sense whose meaning we wish to determine. Suppose that over a given time period the global sensor collects the time-succession of global observations $Z_1, ..., Z_\alpha$ of the time-succession of global states $X_1, ..., X_\alpha$ of the global target, and that the control sensors collect a series $z^*_1, ..., z^*_\alpha$ of global measurements of the states $x^*_1, ..., x^*_\alpha$ of the global sensor (i.e., the states of the individual sensors). Finally, let $u_1, ..., u_\alpha$ be the sequence of global control input vectors that causes the global sensor to "follow" the global target. Consider a fixed time-instant α. At that time-instant, our current knowledge of the state of the global target is completely described by the current global posterior distribution:

$$f_{\alpha \mid \alpha}(X_\alpha, x^*_\alpha \mid Z_1, ..., Z_\alpha, z^*_1, ..., z^*_\alpha, u_1, ..., u_{\alpha-1})$$

Integrating out the control state x^*_α yields the global marginal posterior which describes the statistics of the state of the global target state:

$$f_{\alpha \mid \alpha}(X_\alpha \mid Z_1, ..., Z_\alpha, z^*_1, ..., z^*_\alpha, u_1, ..., u_{\alpha-1})$$

Also, note that by using the global Markov transition density for the global target alone,

$$f_{\alpha+1 \mid \alpha}(X_{\alpha+1} \mid X_\alpha)$$

this information can be extrapolated to time-instant $\alpha+1$:

$$g_{\alpha+1 \mid \alpha}(X_{\alpha+1}) \overset{d}{=} f_{\alpha+1 \mid \alpha}(X_{\alpha+1} \mid Z_1, ..., Z_\alpha, z^*_1, ..., z^*_\alpha, u_1, ..., u_{\alpha-1})$$

This global density comprehensively describes the state of our information concerning the global target at time instant $\alpha+1$, conditioned on the past measurement- and control-history up to time-instant α. On the other hand, the global Markov transition density for the entire system

$$f_{\alpha+1 \mid \alpha}(X_{\alpha+1}, x^*_{\alpha+1} \mid X_\alpha, x^*_\alpha, u_\alpha)$$

can be used to time-update the global posterior at time-instant α to time-instant $\alpha+1$:

$$f_{\alpha+1 \mid \alpha}(X_{\alpha+1}, x^*_{\alpha+1} \mid Z_1, ..., Z_\alpha, z^*_1, ..., z^*_\alpha, u_1, ..., u_{\alpha-1}, u_\alpha)$$

which, again computing the marginal distribution, leads to:

$$g_{\alpha+1 \mid \alpha}(X_{\alpha+1} \mid u_\alpha) \overset{d}{=} f_{\alpha+1 \mid \alpha}(X_{\alpha+1} \mid Z_1, ..., Z_\alpha, z^*_1, ..., z^*_\alpha, u_1, ..., u_\alpha)$$

Finally, suppose that an additional observation $Z_{\alpha+1}$ of the global target is collected. Then we

can compute the updated global marginal posterior

$$g_{\alpha+1\,|\,\alpha+1}(X_{\alpha+1} \mid Z_{\alpha+1}, \mathbf{u}_\alpha) \overset{d}{=} f_{\alpha+1\,|\,\alpha+1}(X_{\alpha+1} \mid Z_1, \ldots, Z_\alpha, Z_{\alpha+1}, z^*_1, \ldots, z^*_\alpha, \mathbf{u}_1, \ldots, \mathbf{u}_{\alpha-1}, \mathbf{u}_\alpha)$$

This global density comprehensively describes the state of our information concerning the global target at time-instant $\alpha+1$:

1) conditioned on the past measurement- and control-history up to time-instant α;

2) conditioned on the control variable $\mathbf{u}_\alpha$ (whose value has yet to be determined); and

3) based on the new global measurement $Z_{\alpha+1}$ of the state of the global target.

If additional information about the state of the global target has become available at time-instant $\alpha+1$, therefore, it must be the result of two things:

- the additional information contained in the global observation $Z_{\alpha+1}$, and

- the additional information resulting from a "good" choice of the global control vector $\mathbf{u}_\alpha$

We are interested only in the latter, since we wish to determine a value of the global control vector $\mathbf{u}_\alpha$ which will have the global sensor "pointing" at the global target as optimally as possible, even *before* we collect a new observation. If we set $Z_{\alpha+1} = \varnothing$ (nothing at all is observed at time-instant $\alpha+1$) then the global density

$$g_{\alpha+1\,|\,\alpha+1}(X_{\alpha+1} \mid \varnothing, \mathbf{u}_\alpha)$$

is a measure of the information at time-instant $\alpha+1$ that is due to "good" sensor allocation via the control-action $\mathbf{u}_\alpha$ alone, independently of any further information that might result from additional measurements. Accordingly, the global control $\mathbf{u}_\alpha$ has been chosen "optimally" if the information contained in

$$g_{\alpha+1\,|\,\alpha+1}(X_{\alpha+1} \mid \varnothing, \mathbf{u}_\alpha)$$

due to the latest control, is as large as possible compared to the information contained in

$$g_{\alpha+1\,|\,\alpha}(X_{\alpha+1})$$

which is due to past history up to time α alone. Or, stated in other terms: The increase in *entropy* (i.e., decrease in information) due to the control should be as small as possible. From this point of view, the density

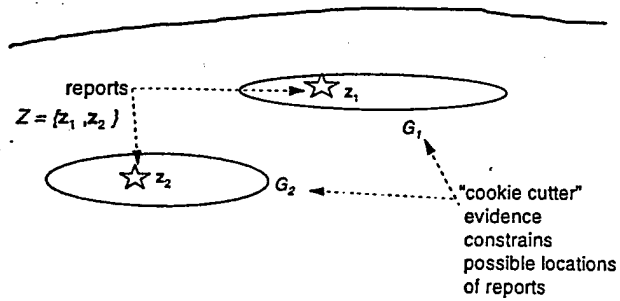$$g_{\alpha+1\,|\,\alpha}(X_{\alpha+1})$$

Figure 1: The statement "target 1 is *NEAR* location A and target 2 is *NEAR* location B" is illustrated when the concept *NEAR* is interpreted as constraining observations within "cookie cutter" régions G_1 and G_2 of locations A,B.
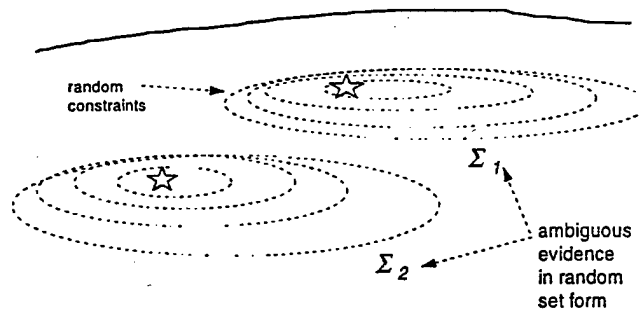

Figure 2: The statement "target 1 is *NEAR* location A and target 2 is *NEAR* location B" illustrated when the concept *NEAR* is interpreted as a variable constraint by *random subsets* Θ_1, Θ_2.

25

is an analog of a target "reference variable" and the density

$$g_{\alpha+1 \mid \alpha+1}(X_{\alpha+1} \mid \varnothing, \mathbf{u}_\alpha)$$

is an analog of a sensor "controlled variable," with entropy being the measure of "closeness" between the two. Accordingly, we need a metric which measures increases in information (as represented in the form of global densities). The *global Kullback-Leibler* information metric has already been defined as

$$K(f; g) \;=\; \int f(X) \, \ln\!\left(\frac{f(X)}{g(X)}\right) \delta X$$

Based on the preceding considerations, we define the following optimality criterion for multisensor, multitarget optimal target-tracking control: Let $f_{\alpha+1 \mid \alpha}(X_{\alpha+1})$ and $f_{\alpha+1 \mid \alpha+1}(X_{\alpha+1} \mid \varnothing, \mathbf{u}_\alpha)$ be as defined previously. Then we define the following cost function:

$$K(\mathbf{u}_1, \dots, \mathbf{u}_M) \;=\; \sum_{\alpha=0}^{M-1} \left[K(f_{\alpha+1 \mid \alpha+1}; f_{\alpha+1 \mid \alpha}) + \mathbf{u}_\alpha{}^T P \mathbf{u}_\alpha \right]$$

We say that a "optimal target-tracker control law" for the multisensor, multitarget allocation problem is a sequence $\mathbf{u}_1, \dots, \mathbf{u}_M$ of control inputs which minimizes this cost function. (For more details see [32]).

g. Data Fusion Using Ambiguous Evidence. In section 2.3 of Chapter 2 of [12] it is shown that many kinds of ambiguous evidence can be modeled as random sets. It is possible to show, as is done in Chapter 7 of [12], that the two sides of data fusion--multisensor, multitarget estimation on the one hand and expert-systems theory on the other--can be fully integrated.

Specifically, one begins by modeling ambiguous observations as *discrete closed random subsets of observation space* and then specifying *measurement models* for ambiguous evidence. Suppose that we have a statement such as

"target 1 is *NEAR* location A and target 2 is *NEAR* location B"

Suppose that *NEAR* in these two cases can be interpreted to mean that target 1 (and therefore observations of target 1) will always be found in "cookie cutter" region G_1 and, likewise, that observations of target 2 will always be found in "cookie cutter" region G_2. Then Figure 1 illustrates the constraint that is imposed by this evidence. In general, however, evidence will consist not of a simple cookie-cutter constraint G_1 but rather of a range of constraints $G_1^{(1)} \subseteq \dots \subseteq G_1^{(d)}$, each being an interpretation of the concept *NEAR* to some degrees $r_1, \dots, r_d$ of likelihood. This evidence can be modeled as a random subset θ_1 of observation space such that $p(\theta_1 = G_1^{(i)}) = r_i$ for all $i = 1, \dots, d$, as illustrated in Figure 2.

Measurement models for ambiguous observations consist of *global measurement densities* which have the form

$$f(Z_1,\ldots,Z_m,\theta_1,\ldots,\theta_{m'} \mid X)$$

where $Z_1,\ldots,Z_m$ denote precise global observations, and where $\theta_1,\ldots,\theta_{m'}$ are discrete random closed subsets of observation space which model ambiguous evidence. From these measurement models one can then define global posterior densities *conditioned on both data and evidence*:

$$f(X \mid Z_1,\ldots,Z_m,\theta_1,\ldots,\theta_{m'})$$

Then one can derive *recursive update equations* for both data and evidence. For example, applying one particular measurement model and suitable independence assumptions, we get

$$f(X \mid Z_1,\ldots,Z_m,\theta_1,\ldots,\theta_{m'}) \propto f(Z_m \mid X)\, f(X \mid Z_1,\ldots,Z_{m-1},\theta_1,\ldots,\theta_{m'})$$

$$f(X \mid Z_1,\ldots,Z_m,\theta_1,\ldots,\theta_{m'}) \propto \beta(\theta_{m'} \mid X)\, f(X \mid Z_1,\ldots,Z_m,\theta_1,\ldots,\theta_{m'-1})$$

where $\beta(\theta \mid X)$ describes the influence of evidence θ. It thereby becomes possible to extend the Bayesian nonlinear filtering equations so that both precise and ambiguous observations can be accommodated into dynamic multisensor, multitarget estimation. For example, if one assumes one possible measurement model for evidence, one gets the following update equation:

$$f_{\alpha+1 \mid \alpha+1}(X_{\alpha+1} \mid Z^{(\alpha)}, \theta^{(\alpha+1)}) = \frac{\beta(\theta_{\alpha+1} \mid X_{\alpha+1})\, f_{\alpha+1 \mid \alpha}(X_{\alpha+1} \mid Z^{(\alpha)}, \theta^{(\alpha)})}{\int \beta(\theta_{\alpha+1} \mid Y_{\alpha+1})\, f_{\alpha+1 \mid \alpha}(Y_{\alpha+1} \mid Z^{(\alpha)}, \theta^{(\alpha)})\, \delta Y_{\alpha+1}}$$

h. A Bayesian Characterization of Rules of Evidential Combination. If one specific measurement model for ambiguous evidence is assumed--the so-called "data-dependent" model (see Chapter 7 of [12])--then posterior distributions will have the property that

$$f(X \mid Z_1,\ldots,Z_m,\theta_1,\ldots,\theta_{m'}) = f(X \mid Z_1,\ldots,Z_m,\theta_1 \cap \ldots \cap \theta_{m'})$$

In other words, the random-set intersection operator '$\cap$' may be interpreted as a means of fusing multiple pieces of ambiguous evidence in such a way that posteriors conditioned on the fused evidence are identical to posteriors conditioned on the individual evidence and computed using Bayes' rule alone. Thus, for example, suppose that

$$\theta_1 = \Sigma_A(f) = \{u \in U \mid A \leq f(u)\}$$

$$\theta_2 = \Sigma_A(g) = \{u \in U \mid A \leq g(u)\}$$

where f,g are two fuzzy membership functions on U and A is a uniformly distributed random number on $[0,1]$. Let '$\wedge$' denote the Zadeh "min" fuzzy *AND* operation and *define* the posteriors $f_{\Gamma \mid \Sigma}(X \mid Z, f \wedge g)$ and $f_{\Gamma \mid \Sigma}(X \mid Z, f, g)$ by

$$f_{\Gamma \mid \Sigma}(X \mid Z, f, g) = f(X \mid Z, \Sigma_A(f), \Sigma_A(g))$$

27

$$f_{\Gamma\mid\Sigma}(X \mid Z, f \wedge g) = f(X \mid \Sigma_A(f \wedge g))$$

We know from section 2.3.2 of Chapter 2 of **[12]** that

$$\Sigma_A(f) \cap \Sigma_A(g) = \Sigma_A(f \wedge g)$$

and thus that

$$f_{\Gamma\mid\Sigma}(X \mid Z, f, g) = f(X \mid Z, \Sigma_A(f), \Sigma_A(g)) = f(X \mid Z, \Sigma_A(f) \cap \Sigma_A(g))$$

$$= f(X \mid Z, \Sigma_A(f \wedge g)) = f(X \mid Z, f \wedge g)$$

and so

$$f_{\Gamma\mid\Sigma}(X \mid Z, f, g) = f(X \mid Z, f \wedge g)$$

That is: The fuzzy *AND* is a means of fusing fuzzy evidence in such a way that posteriors conditioned on the fused evidence are identical to posteriors conditioned on the fuzzy evidence individually and computed using Bayes' rule alone. Thus fuzzy logic is entirely consistent with Bayesian probability--*provided that it is first represented in random set form*, and *provided that we use a specific measurement model for ambiguous observations*. Similar observations apply to *any* rule of evidential combination which bears a homomorphic relationship with the random set intersection operator.

4-B-g. Possible Objections to "Finite-Set Statistics". A number of objections to the use of random set theory in data fusion can be anticipated. We address some of these in turn.

Isn't It Just "Bookkeeping"? The reformulation of multisource, multitarget problems as single-sensor, single-target problems is not just a mathematical "bookkeeping" device. Generally speaking, *any group of targets observed by imperfect sensors must be analyzed as a single indivisible entity rather than as a collection of unrelated individuals.* When measurement uncertainties are large in comparison to target separations there will always be a significant likelihood that any given measurement was generated by any given target. This means that *every* measurement can be associated, partially or in some degree of proportion, to *every* target. The more irresolvable the targets are, the more our estimates of them will be statistically correlated and thus the more that they will seem as though they are a *single* target. Observations can no longer be regarded as separate entities generated by individual targets but rather as collective phenomena *generated by the entire multitarget system.* This remains true even when target separations are large in comparison to sensor uncertainties. Though in this case the likelihood is very small that a given observation was generated by any other target than the one it is intuitively associated with, nevertheless this likelihood is nonvanishing. Thus the intuitive "this observation goes with that track" perspective is only an ideal limiting case of the more general multitarget problem. The random set approach inherently models--and forces one to take account of--the organic/collective nature of multitarget systems.

Why Not Vectors or Point Processes? Other skeptics might ask: *Why not simply use vector models or point process models?* It could be objected, for example, that what we have called the "global density" $f_\Sigma(Z)$ of a finite random subset Σ is just a new name and notation for the Janossy densities $j_n(z_1,\ldots,z_n)$, $n \geq 0$ [5], pp. 122-123) of the corresponding simple finite point process defined by $N_\Sigma(S) = |\Sigma \cap S|$ for all measurable S. (Point processes have been investigated as a basis for multitarget tracking in [59, 65], for example.) In response to possible such objections we offer the following responses.

First, vector approaches *encourage carelessness in regard to basic questions*. For example, to apply the theorems of conventional estimation theory one must clearly identify a measurement space and a state space and specify their topological and metrical properties. Wald's proof of the consistency of the maximal likelihood and maximum *a posteriori* estimators, for example, assumes that state space is a *metric space* satisfying certain properties. As another instance suppose that we want to determine whether small deviations in the input data to a data fusion algorithm can result in large deviations in output data. To answer this question one must first have some idea of what *distance* means in both measurement space and state space. The standard Euclidean metric is clearly not adequate: If we represent an observation set $\{z_1, z_2\}$ as a vector (z_1, z_2) then $\| (z_1, z_2) - (z_2, z_1) \| \neq 0$ even though the order of measurements should not matter. Likewise one might ask, What is the distance between (z_1, z_2) and z_3? Whereas both finite set theory and point process theory have rigorous metrical concepts, attempts to define metrics for vector models can quickly degenerate into *ad hoc* invention.

More generally, the use of vector models has resulted in piecemeal solutions to information fusion problems (most typically, the assumption that the number of targets is known *a priori*). Lastly, any attempt to incorporate expert systems theory into the vector approach results in extremely awkward attempts to make vectors behave as though they were finite sets.

Second, the random set approach is *explicitly geometric* in that the random variates in question are actual sets of observations--rather than, say, abstract integer-valued measures.

Third, systematic adherence to a random set perspective results in a series of direct parallels between single-sensor, single-target statistics and multisensor, multitarget statistics which results in a methodology for information fusion that is nearly identical in general behavior to the "Statistics 101" formalism with which engineering practitioners and theorists are already familiar. (The elements of the random set approach are simple enough that they could be taught at the junior and senior undergraduate levels. Random measure theory, by way of contrast, is unlikely to ever find its way into the undergraduate curriculum even in mathematics departments.) More importantly, it leads to a systematic approach to solving information fusion problems that allows standard single-sensor, single-target statistical techniques to be directly generalized to the multisensor, multitarget case.

Fourth, because the random set approach provides a systematic foundation for both expert systems theory *and* multisensor, multitarget estimation (see section 2.5.6 of Chapter 2 of [12]), it permits a systematic and mathematically rigorous integration of these two quite different

aspects of information fusion--a question left unaddressed by either the vector or point-process models.

Fifth, an analogous situation holds in the case of random subsets of $\mathbb{R}^n$ which are *convex* and *bounded*. Given a bounded convex subset $T \subseteq \mathbb{R}^n$, the *support function* of T is defined by

$$s_T(\mathbf{e}) = sup_{\mathbf{x} \in T} \langle \mathbf{e}, \mathbf{x} \rangle$$

for all vectors $\mathbf{e}$ on the unit hypersphere in $\mathbb{R}^n$, where '$\langle -,- \rangle$' denotes the inner product on $\mathbb{R}^n$. The assignment $\Sigma \to s_\Sigma$ establishes a very faithful embedding of random bounded convex sets Σ into the random functions on the unit hypersphere, in the sense that it encodes the behavior of bounded convex sets into vector mathematics [27, 46]. Nevertheless, it does not follow that the theory of random bounded convex subsets is a special case of random function theory. Rather, random functions provide a useful tool for studying the behavior of random bounded convex sets. In like manner, finite point processes are best understood as specific--and by no means the only or the most useful--representations of random finite subsets as elements of some abstract vector space.

4-B-h. Computational techniques analysis. A survey of potentially applicable computational and approximation techniques was accomplished under the current contract. The techniques thus far identified as suitable for further investigation are as follows.

Approximate Computation of Permanents of Matrices. The most numerically intensive computation involved in random set formulations of multitarget data fusion problems consists of evaluating or approximating combinatorial sums of the form

$$\sum_{A \subseteq Q} perm(Q)$$

where $Q = \{Q_{i,j}\}_{i=1,...,n; j=1,...,m}$ is an arbitrary real-valued matrix, where the summation is taken over all *square* submatrices $A = \{A_{i,j}\}_{i,j=1,...,e}$ of Q, and where

$$perm(A) = \sum_\sigma A_{1,\sigma 1} \cdots A_{e,\sigma e}$$

is the permanent [2] of A (summation taken over all permutations σ on the numbers $1,...,e$). The reason for this is as follows. Let $\mathbf{z}_1,...,\mathbf{z}_m$ be a scan of observations, let $\mathbf{x}_1,...,\mathbf{x}_t$ be a list of target parameters with $m < n$, and define

$$Q_{i,j} = f(\mathbf{z}_i \mid \mathbf{x}_j)$$

with $i = 1,...,m; j = 1,...,n$. Then, assuming independence of measurements, exact computation of global densities requires computation of all permanents of all square submatrices of matrices of the form Q.

Permanents can be computed exactly with computational complexity 2^e [50]. Permanents can

30

also be approximated in a far more computationally efficient manner, however, because of the fact that they share some of the same properties as determinants. For example, if Q can be written as a block-diagonal matrix whose blocks are the square submatrices $Q_1, ..., Q_e$ then $perm(Q) = perm(Q_1) \cdots perm(Q_e)$. Also, permanents can be expanded by rows or by columns, and are unchanged by interchanges of rows or interchanges of columns. Approximate computation of permanents is achieved by "sparsifying" confusion matrices--i.e., by zeroing out any entries which are excessively small compared to other entries (that is, any low-likelihood associations are ignored). A number of other methods for approximating the values of permanents have been investigated [23, 54].

Approximation by Gaussian Sums. The Gaussian sum approach for approximating time-propagated general Bayesian posterior distributions was introduced over twenty years ago by Aspach and Sorenson [62, 63]. The basic idea is to approximate the various densities involved in Bayesian recursive nonlinear filtering--the prior, state-transition, and measurement-model densities--as Gaussian mixtures (finite sums of Gaussian distributions). Gaussian sums are closed under the recursive filtering operations (integration and time prediction; division and Bayesian update). As a result, the time-evolved posterior densities are also Gaussian sums, though the number of terms in these sums increases exponentially with time and thus pruning of terms is necessary.

If the underlying sensors and target motion models are Gaussian, it is easy to demonstrate that the corresponding multitarget posterior distributions (global posteriors) are themselves Gaussian sums. Accordingly, in the purely Gaussian case the multisensor, multitarget problem is naturally suited for application of Gaussian sum approximation techniques. More generally, suppose that the prior, measurement-model, and motion-model densities for individual sensors and targets, respectively, are approximated as Gaussian sums. Then the corresponding global densities are also approximated as Gaussian sums. Since the global recursive nonlinear filtering equations (discussed earlier) are closed with respect to Gaussian sums, it follows that approximation techniques of this kind may prove useful in more general data fusion problems as well.

Asymptotic Approximations of Integrals. The simplest of asymptotic approximations is the saddle-point approximation (also known as Lagrange's method of integration). It is a technique for asymptotically approximating integrals of suitably well-behaved functions ([55], pp. 88-90 and [20], pp. 30-37). The simplest and most widely known application of the method is to deriving Sterling's approximation formula for factorials. That is, from the general formula

$$z! = \int_0^\infty t^z e^{-t} dt = \int_0^\infty e^{z \ln t - t} dt$$

for the factorial function, one notes that, as a function of t, the integrand is largest when

$$\frac{\partial}{\partial t}\, (z \ln t - t) \;=\; 0$$

i.e. at the saddle point $t = z$. Expanding $z \ln t - t$ as a function of t in a Taylor's series about $t = z$ leads to

$$z \ln t - t \;=\; (z \ln z - z) + 0\,(t-z) - \frac{1}{2z}\,(t-z)^2 + \ldots$$

and thus

$$z! \;=\; \int_0^\infty e^{z \ln t - t}\, dt \;\approx\; e^{-z} z^z \int_0^\infty \exp\!\left(-\frac{1}{2z}\,(t-z)^2\right) dt \;=\; e^{-z} z^z \sqrt{2 \pi z}$$

since the saddle point value will dominate the integral for large z.

Lagrange's method has been employed in statistics since the 1960s to approximate the values of various statistical quantities defined in terms of definite integrals (see [6, 58]), assuming a large number of data samples. It would be desirable to determine whether or not Lagrange's method can be generalized to approximate quantities defined in terms of *set integrals*. A broader class of asymptotic approximation methods known as large-deviation techniques (see [3, 7]) are also of interest.

Computational Statistical Mechanics. Set integrals occur regularly in the statistical theory of gases and liquids (see [19], pp. 234, Equ. 37.4; p. 266, Equ. 40.28) though they are never explicitly identified as such in that context. This should not be surprising since there are many formal mathematical similarities between multitarget systems with multiple target types, on the one hand, and ensembles of molecules belonging to multiple molecular species, on the other. Because of this, there is reason to believe that the physics community has developed approximate computational techniques which might be applicable to multisource, multitarget information fusion problems (see, for example [16]).

Lockheed Martin has already had some success [24] with one such statistical approximation technique known in the physics community as "mean-field approximation" [52]. This approach provides a means of approximating combinatorial sums by first approximating them as integrals and then applying Laplace's integration method. To illustrate the approximation of a relatively simple combinatorial sum using this technique, we apply it to the problem of computing the permanent of an $n \times n$ matrix $Q = \{Q_{j,k}\}_{j,k=1,\ldots,n}$:

$$perm(Q) \;=\; \sum_\sigma Q_{1,\sigma 1} \cdots Q_{n,\sigma n}$$

where the sum is taken over all permutations σ on the numbers $1,\ldots,n$. Let $\Pi(n)$ denote the

set of all $n \times n$ permutation matrices, i.e. the set of all $n \times n$ matrices $A = \{a_{j,k}\}_{j,k=1,\ldots,n}$ where $a_{j,k}$ are nonnegative integers such that $\Sigma_{j=1}^{n} a_{j,k} = 1$ for all $k = 1,\ldots,n$ and $\Sigma_{k=1}^{n} a_{j,k} = 1$ for all $j = 1,\ldots,n$. (In other words, each row and each column of A contains at most one nonzero entry, which must in turn be 1.) Then we can write

$$perm(Q) \quad = \quad \sum_{A \in \Pi(n)} Q_{1,1}^{a_{1,1}} \cdots Q_{j,k}^{a_{j,k}} \cdots Q_{n,n}^{a_{n,n}} \quad = \quad \sum_{A \in \Pi(n)} \prod_{j,k=1}^{n} Q_{j,k}^{a_{j,k}}$$

Now, let

$$D(A) \quad = \quad \prod_{k=1}^{n} \delta_0\left(-1 + \sum_{j=1}^{n} a_{j,k}\right)$$

where $\delta_0(x)$ is the Kronecker delta function defined by $\delta_0(x) = 1$ if $x = 0$ and $\delta_0(x) = 0$ otherwise. Then $D(A) = 0,1$ and $D(A) = 1$ if and only if $\Sigma_{j=1}^{n} a_{j,k} = 1$ for all $k = 1,\ldots,n$. Consequently, we can write

$$perm(Q) \quad = \quad \sum_{A \in \Pi^*(n)} D(A) \prod_{j,k=1}^{n} Q_{j,k}^{a_{j,k}}$$

where $\Pi^*(n)$ denotes the set of all $n \times n$ matrices A such that $a_{j,k}$ are nonnegative integers and such that $\Sigma_{k=1}^{n} a_{j,k} = 1$ for all $j = 1,\ldots,n$.

Next, approximate δ_0 as a limit of exponential functions,

$$\delta_0(x) \; = \; \lim_{N \to \infty} e^{-\frac{1}{2}Nx^2} \; = \; \lim_{N \to \infty} \int_{-\infty}^{\infty} e^{ixy} \, d\tilde{y} \,, \qquad d\tilde{y} \; = \; \frac{1}{\sqrt{2\pi N}} \, e^{-y^2/2N} \, dy$$

and substitute the integral expression into $D(A)$. Then after a little algebra we get:

$$perm(Q) \quad = \quad \lim_{N \to \infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-i\Sigma_{k=1}^{n} y_k} \sum_{A \in \Pi^*(n)} e^{\Sigma_{j,k=1}^{n} a_{j,k}(iy_k + \ln Q_{j,k})} \, d\tilde{y}_1 \cdots d\tilde{y}_n$$

Now, let us be given any family of numbers $m(A) = \Sigma_{j,k=1}^{n} a_{j,k} b_{j,k}$ indexed by the matrices $A \in \Pi^*(n)$, where $B = \{b_{j,k}\}_{j,k=1,\ldots,n}$ is any fixed real-valued $n \times n$ matrix. Then note that

$$m(A) \; = \; (a_{1,1}b_{1,1} + \ldots + a_{1,n}b_{1,n}) + (a_{2,1}b_{2,1} + \ldots + a_{2,n}b_{2,n}) + \ldots + (a_{n,1}b_{n,1} + \ldots + a_{n,n}b_{n,n})$$

where $a_{j,1} + \ldots + a_{j,n} = 1$ for every $j = 1,\ldots,n$ is identical to the family

$$m(k_1,...,k_n) = b_{1,k_1} + ... + b_{n,k_n} = \sum_{j=1}^{n} b_{j,k_j}$$

indexed by the integers $k_1,...,k_n = 1,...,n$. Therefore we can write:

$$perm(Q) = \lim_{N \to \infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-i\sum_{k=1}^{n} y_k} \sum_{k_1,...,k_n=1}^{n} e^{\sum_{j=1}^{n}(iy_{k_j} + \ln Q_{j,k_j})} d\tilde{y}_1 \cdots d\tilde{y}_n$$

(In other words, the combinatorial sum has been replaced by a non-combinatorial sum.) More algebra and a substitution of variables leads to the following complex line integral:

$$perm(Q) = \lim_{N \to \infty} \frac{1}{\sqrt{2^n \pi^n N^n}} \frac{1}{i} \int_{-\infty i}^{\infty i} \cdots \frac{1}{i} \int_{-\infty i}^{\infty i} e^{E_N(s_1,...,s_n)} ds_1 \cdots ds_n$$

where

$$E_N(s_1,...,s_n) = \sum_{k=1}^{n} \left(\frac{s_k^2}{2N} - s_k \right) + \sum_{j=1}^{n} \ln \left(\sum_{k=1}^{n} Q_{j,k} e^{s_k} \right)$$

The complex line integrals can then be approximated using the saddle point method described earlier. The mean-field approximation has the effect of reducing the computational complexity of *perm(Q)*, which is of order 2^n, to an approximation of order n^3. For more details see [24].

4-C. LIST OF PUBLICATIONS AND TECHNICAL REPORTS

The following publications were generated under this contract.

4-C-a. Books (Work Supported by USARO)

1. I.R. Goodman, R.P.S. Mahler, and H.T. Nguyen (1997) *Mathematics of Data Fusion*, Kluwer Academic Publishers

2. J. Goutsias, R. Mahler, and H.T. Nguyen, eds. (1997) *Theory and Applications of Random Sets*, Springer-Verlag

4-C-b. Conference and Other Papers Published in Books (Work Supported by USARO)

1. R. Mahler (1997) "A Theoretical Unification of Knowledge-Based Systems With Multisensor, Multitarget Estimation," in P. Wang, ed., *Advances in Machine Intelligence and Soft Computing, Vol. IV*, Duke University Dept. of Elec. Eng., to appear.

2. R. Mahler (1996/1994) "A Unified Foundation for Data Fusion," in F.A. Sadjadi, ed., *Selected Papers on Sensor and Data Fusion*, SPIE Vol. MS-124, SPIE Optical Engineering Press, Bellingham WA, pp. 325-345; reprinted from *Proc. 7th Joint Service Data Fusion Symposium,* Vol. I Part 1 (Unclassified), Oct. 25-28 1994, Johns Hopkins Applied Physics Laboratory, Laurel MD, Naval Air Development Center, Warminster PA, pp. 153-173

3. R. Mahler (1995) "Finite-set statistics with application to data fusion," in A. Friedman, ed., *Mathematics in Industrial Problems, Part 7*, Springer-Verlag, pp. 198-206

4. R. Mahler (1994) "Systematic data fusion using the theory of conditional random sets," in A. Friedman, ed., *Mathematics in Industrial Problems, Part 6*, pp. 156-165

4-C-c. Conference Papers Published in Proceedings (Work Supported by USARO)

1. R. Mahler (1998) "Global posterior densities for sensor management," *Proc. 1998 SPIE AeroSense Conference*, April 13-17 1998, Orlando, to appear

2. R. Mahler (1998) "Information for fusion management and performance estimation," *Proc. 1998 SPIE AeroSense Conference*, April 13-17 1998, Orlando, to appear

3 R. Mahler (1998) "Multisource, multitarget filtering: A unified approach," *Proc. 1998 SPIE AeroSense Conference*, April 13-17 1998, Orlando, to appear

4. R. Mahler (1997) "Decisions and Data Fusion," *Proc. 1997 National Symposium on Sensor and Data Fusion*, April 14-17 1997, M.I.T. Lincoln Laboratories, to appear

5. R. Mahler (1997) "Measurement models for ambiguous evidence using conditional random sets," *Proc. 1997 SPIE Aerosense Conf.*, April 21-25 1997, Orlando, to appear

6. R. Mahler (1996) "Global Optimal Sensor Allocation," *Proceedings of the Ninth National Symposium on Sensor Fusion*, Vol. I (Unclassified), Mar. 12-14 1996, Naval Postgraduate School, Monterey CA, pp. 347-366

7. R. Mahler (1996) "Unified data fusion: fuzzy logic, evidence, and rules," in I. Kadar and V. Libby, eds., *Signal Processing, Sensor Fusion, and Target Recognition V,* SPIE Vol. 2755, SPIE Optical Engineering Press, Bellingham WA, pp. 226-237

8. R. Mahler (1995) "Information Theory and Data Fusion," *Proceedings of the Eighth National Symposium on Sensor Fusion*, Vol. I (Unclassified), Dallas TX, March 15-17 1995, ERIM, Ann Arbor, pp. 279-292

9. R. Mahler (1995) Nonadditive probability, finite-set statistics, and information fusion" (*Invited Paper*), *Proc. 34th IEEE Conf. on Decision and Control*, New Orleans, Dec.

1995, pp. 1947-1952

10. R. Mahler (1995) "Unified nonparametric data fusion" (*Invited Paper*), in I. Kadar and V. Libby, eds., *Signal Processing, Sensor Fusion, and Target Recognition IV*, SPIE Vol. 2484, SPIE Optical Engineering Press, Bellingham WA, pp. 66-74

11. R. Mahler (1994) "Global Integrated Data Fusion" (*Invited Paper*), *Proceedings of the Seventh National Symposium on Sensor Fusion*, Vol. I (Unclassified), Sandia National Laboratories, Albuquerque NM, March 16-18 1994, ERIM, Ann Arbor, pp. 187-199

12. R. Mahler (1994), "The random set approach to data fusion," in F.A. Sadjadi, ed., *Automatic Object Recognition IV*, SPIE Vol. 2234, SPIE Optical Engineering Press, Bellingham WA, 1994, pp. 287-295

4-C-d. Journal Articles Published During Performance Period (Not Supported by ARO)

1. D. Fixsen and R. Mahler (1997) "The Modified Dempster-Shafer Approach to Classification," *IEEE Trans. Systems, Man and Cybernetics, Part A*, vol. 27 no. 1, pp. 96-104

2. R. Mahler (1996) "Combining Ambiguous Evidence With Respect to Ambiguous *a priori* Knowledge, I: Boolean Logic," *IEEE Trans. Systems, Man and Cybernetics, Part A*, vol. 26, pp. 27-41

3. R. Mahler (1996) "Representing Rules as Random Sets: Statistical Correlations Between Rules," *Information Sciences*, vol. 88, pp. 47-68

4. R. Mahler (1996) "Representing Rules as Random Sets, II: Iterated Rules," *Intn'l Jour. Intell. Sys.*, vol. 11, pp. 583-610

5. R. Mahler (1995) "Combining Ambiguous Evidence With Respect to Ambiguous *a priori* Knowledge, II: Fuzzy Logic," *Fuzzy Sets and Systems*, vol. 75, pp. 319-354

4-C-e. Conference Papers Published During Performance Period (Not Supported by ARO)

1. J. Hatlestad, R. Mahler, T.C. Poling, R. Allen, R. Myre, J. Warner, M. Hatch, E. Jahn, and J. Kaina (1997) "ASW/ASUW Parametric ESM/ACINT Multitarget Classifier-Tracker," *Proc. 1997 Combat Identification Systems Conf.*, to appear

4-D. LIST OF ALL PARTICIPATING SCIENTIFIC PERSONNEL

● **Ronald P.S. Mahler**   **POSITION TITLE:** Staff Scientist

Education

Institution & Location	Degree	Year	Field of Study
U. of Chicago, Chicago IL	B.A.	1970	Mathematics
U. of Minnesota, Mpls MN	B.E.E.	1980	Electrical Engineering
Brandeis U., Waltham MA	Ph.D.	1974	· Mathematics

Dr. Mahler served as assistant professor of mathematics in the School of Mathematics of the University of Minnesota, Minneapolis MN, from 1974-1979. He has been employed at the Eagan MN facility of Lockheed Martin since 1980. He is currently on the program committee, and is a session chair, of the data fusion conference of the SPIE Aerosense conference. He was also principal organizer, co-chair, and co-editor for a Workshop on Applications and Theory of Random Sets, held at the Institute for Mathematics and Its Applications (Minneapolis), jointly sponsored by ONR, ARO, and Lockheed Martin. The proceedings of this workshop are to appear in hardcover (Springer-Verlag) in 1997.

● **Paul Leavitt**   **POSITION TITLE:** Software Analyst

Education

Institution & Location	Degree	Year	Field of Study
	BSEE	1985	Electrical Engineering
	BSPh	1979	Pharmacy

Mr. Leavitt has an extensive background in Real Time Computer software and hardware design. He has extensive experience in real-world applications, including: Air Superiority Fighter computational requirements analysis for the ATF Program; 1750A computer design and microcode design for the ATF and B2 Programs. He has Integrated GPS Interferometry with Inertial Navigation components to provide a low cost Agricultural Navigation System (AGNAV). He has an excellent background in data analysis and algorithm design and implementation in such areas as GPS Carrier Phase ambiguity resolution and Attitude determination, Kalman filtering. He has proficiency in communications, such as Spread Spectrum Data Links, T1, E1, and proprietary algorithms used in communications programs such as ABCCC. His current tasks are related to implementation of proprietary algorithms on the Mercury parallel computer architecture.

5. REPORT OF INVENTIONS

None

6. BIBLIOGRAPHY

1. R.T. Antony (1995) *Principles of Data Fusion Automation*, Artech House, Boston

2. R.A. Brualdi and H.J. Ryser (1991) *Combinatorial Matrix Theory*, Cambridge University Press, Chapter 7

3. J.A. Bucklew (1990) *Large Deviation Techniques in Decision, Simulation, and Estimation*, John Wiley and Sons

4. C.-Y. Chong, S. Mori, and K.C. Chang (1990) "Distributed Multitarget, Multisensor Tracking," in Y. Bar-Shalom, ed., *Multitarget-Multisensor Tracking*, Artech House, pp. 247-295

5. D.J. Daley and D. Vere-Jorre, *An Introduction to the Theory of Point Processes*, Springer-Verlag 1988

6. H.E. Daniels, "Saddlepoint Approximations in Statistics," *Annals of Math. Stat.*, vol. 25, pp. 631-650

7. R.S. Ellis (1985) *Entropy, Large Deviations, and Statistical Mechanics*, Springer-Verlag

8. D. Fixsen and R. Mahler (1997) "The Modified Dempster-Shafer Approach to Classification," *IEEE Trans. Systems, Man and Cybernetics, Part A*, vol. 27 no. 1, pp. 96-104

9. I.R. Goodman (1994) "A New Characterization of Fuzzy Logic Operators Producing Homomorphism-Like Relations With One-Point Coverages of Random Sets," in P.P. Wang, ed., *Advances in Fuzzy Theory and Technology*, Vol. II, pp. 131-157

10. I.R. Goodman (1983) "A unified approach to modeling and combining of evidence through random set theory, *Proc. 6th MIT/ONR Workshop on C3 Sysems*, Cambridge MA, pp. 42-47

11. I.R. Goodman (1982) "Fuzzy sets as equivalence classes of random sets," in R. Yager, ed., *Fuzzy Sets and Possibility Theory*, Permagon, pp. 327-343

12. I.R. Goodman, R.P.S. Mahler, and H.T. Nguyen (1997) *Mathematics of Data Fusion*, Kluwer Academic Publishers, 1997

13. I.R. Goodman and H.T. Nguyen (1985) *Uncertainty Models for Knowledge Based Systems*, North-Holland, Amsterdam

14. J. Goutsias, R. Mahler, and H.T. Nguyen, eds. (1997) *Theory and Applications of Random Sets*, Springer-Verlag, 1997

15. M. Grabisch, H.T. Nguyen, and E.A. Walker (1995) *Fundamentals of Uncertainty Calculi With Applications to Fuzzy Inference*, Kluwer Academic Publishers

16. L. Greengard (1994) "Fast Algorithms for Classical Physics," *Science*, vol. 265, pp. 909-914

17. D.L. Hall, *Mathematical Techniques in Multisensor Data Fusion*, Artech House, 1992

18. K. Hestir, H.T. Nguyen, and G.S. Rogers (1991) "A random set formalism for evidential reasoning," in I.R. Goodman, M.M. Gupta, H.T. Nguyen and G.S. Rogers, editors, *Conditional Logic in Expert Systems*, North-Holland, Amsterdam, pp. 309-344

19. T.L. Hill (1987) *Statistical Mechanics: Principles and Practical Applications*, Dover Publications

20. E.J. Hinch (1991) *Perturbation Methods*, Cambridge University Press

21. S.J. Julier and J.K. Uhlmann (1997) "A Practical Real-Time Algorithm for General Nonlinear Filtering," *Proc. 1997 SPIE AeroSense Conference*, April 20-25 1997, Orlando FL, Vol. 3068, to appear

22. S.J. Julier, J.K. Uhlmann, and H.F. Durrant-Whyte (1995) "A New Approach for Filtering Nonlinear SYstem," *Proc. Amer. Contr. Conf.*, Seattle WA, pp. 1628-1632

23. N. Karmarkar, R. Karp, R. Lipton, L. Lovás, and M. Luby (1993) "A Monte-Carlo Algorithm for Estimating the Permanent," *SIAM J. Comput.*, vol. 22 no. 2, pp. 284-293

24. K. Kastella (1995) "Event-Averaged Maximum Likelihood Estimation and Mean-Field Theory in Multitarget Tracking," *IEEE Trans. Aut. Contr.*, vol. 40 no. 6, pp. 1070-1074

25. D.G. Kendall (1963) "Foundations of a theory of random sets," in E.F. Harding and D.G. Kendall, editors, *Stochastic Geometry*, J. Wiley, New York, pp. 322-376

26. R. Kruse, E. Schwencke, and J. Heinsohn (1991) *Uncertainty and Vagueness in Knowledge-Based Systems*, Springer-Verlag

27. N.N. Lyshenko, "Statistics of Random Compact Sets in Euclidean Space," *Journal of Soviet Mathematics*, Vol. 21 1983, Plenum Publishing, pp. 76-92

28. R. Mahler (1997) "Decisions and Data Fusion," *Proc. 1997 National Symposium on Sensor and Data Fusion*, April 14-17 1997, M.I.T. Lincoln Laboratories, to appear

29. R. Mahler (1997) "Measurement models for ambiguous evidence using conditional random sets," *Proc. 1997 SPIE Aerosense Conf.*, April 21-25 1997, Orlando, to appear

30.   R. Mahler (1997) "A Theoretical Unification of Knowledge-Based Systems With Multisensor, Multitarget Estimation," in P. Wang, ed., *Advances in Machine Intelligence and Soft Computing, Vol. IV*, Duke University Dept. of Elec. Eng., to appear.

31. R. Mahler (1996) "Combining Ambiguous Evidence With Respect to Ambiguous *a priori* Knowledge, I: Boolean Logic," *IEEE Trans. Systems, Man and Cybernetics, Part A*, vol. 26, pp. 27-41

32. R. Mahler (1996) "Global Optimal Sensor Allocation," *Proceedings of the Ninth National Symposium on Sensor Fusion*, Vol. I (Unclassified), Mar. 12-14 1996, Naval Postgraduate School, Monterey CA, pp. 347-366

33. R. Mahler (1996) "Representing Rules as Random Sets: Statistical Correlations Between Rules," *Information Sciences*, vol. 88, pp. 47-68

34. R. Mahler (1996) "Representing Rules as Random Sets, II: Iterated Rules," *Intn'l Jour. Intell. Sys.*, vol. 11, pp. 583-610

35. R. Mahler (1996) "Unified data fusion: fuzzy logic, evidence, and rules," in I. Kadar and V. Libby, eds., *Signal Processing, Sensor Fusion, and Target Recognition V*, SPIE Vol. 2755, SPIE Optical Engineering Press, Bellingham WA, pp. 226-237

36. R. Mahler (1996/1994) "A Unified Foundation for Data Fusion," in F.A. Sadjadi, ed., *Selected Papers on Sensor and Data Fusion*, SPIE Vol. MS-124, SPIE Optical Engineering Press, Bellingham WA, pp. 325-345; reprinted from *Proc. 7th Joint Service Data Fusion Symposium*, Vol. I Part 1 (Unclassified), Oct. 25-28 1994, Johns Hopkins Applied Physics Laboratory, Laurel MD, Naval Air Development Center, Warminster PA, pp. 153-173

37. R. Mahler (1995) "Combining Ambiguous Evidence With Respect to Ambiguous *a priori* Knowledge, II: Fuzzy Logic," *Fuzzy Sets and Systems*, vol. 75, pp. 319-354

38. R. Mahler (1995) "Finite-set statistics with application to data fusion," in A. Friedman, ed., *Mathematics in Industrial Problems, Part 7*, Springer-Verlag, pp. 198-206

39. R. Mahler (1995) "Information Theory and Data Fusion," *Proceedings of the Eighth National Symposium on Sensor Fusion*, Vol. I (Unclassified), Dallas TX, March 15-17 1995, ERIM, Ann Arbor, pp. 279-292

40. R. Mahler (1995) Nonadditive probability, finite-set statistics, and information fusion" (*Invited Paper*), *Proc. 34th IEEE Conf. on Decision and Control*, New Orleans, Dec. 1995, pp. 1947-1952

41. R. Mahler (1995) "Unified nonparametric data fusion" (*Invited Paper*), in I. Kadar and V. Libby, eds., *Signal Processing, Sensor Fusion, and Target Recognition IV*, SPIE Vol. 2484,

SPIE Optical Engineering Press, Bellingham WA, pp. 66-74

42. R. Mahler (1994) "Global Integrated Data Fusion" (*Invited Paper*), *Proceedings of the Seventh National Symposium on Sensor Fusion*, Vol. I (Unclassified), Sandia National Laboratories, Albuquerque NM, March 16-18 1994, ERIM, Ann Arbor, pp. 187-199

43. R. Mahler (1994), "The random set approach to data fusion," in F.A. Sadjadi, ed., *Automatic Object Recognition IV*, SPIE Vol. 2234, SPIE Optical Engineering Press, Bellingham WA, 1994, pp. 287-295

44. R. Mahler (1994) "Systematic data fusion using the theory of conditional random sets," in A. Friedman, ed., *Mathematics in Industrial Problems, Part 6*, pp. 156-165

45. G. Matheron (1975) *Random Sets and Integral Geometry*, J. Wiley, New York

46. I.S. Molchanov (1993) *Limit Theorems for Unions of Random Closed Sets*, Springer-Verlag Lecture Notes in Mathematics No. 1561

47. S. Mori, C.-Y. Chong, E. Tse, and R.P. Wishner (1984) "Multitarget Multisensor Tracking Problems, Part I: A General Solution and a Unified View on Bayesian Approaches," Revised Version, Technical Report TR-1048-01, Advanced Information and Decision Systems, Inc., Mountain View CA, August 1984. My thanks to Dr. Mori for making this report available to me (Dr. Shozo Mori, personal communication, Feb. 28 1995)

48. S. Mori, C.-Y. Chong, E. Tse, and R.P. Wishner (1986) "Tracking and Classifying Multiple Targets Without *A Priori* Identification," *IEEE Transactions on Automatic Control*, Vol. AC-31 No. 5, pp. 401-409

49. H.T. Nguyen (1978) "On random sets and belief functions," *Journal of Mathematical Analysis and Applications*, Vol. 65, pp. 531-542

50. S.D. O'Neil and M.F. Bridgland (1991) "Fast Algorithms for Joint Probabilistic Data Association," *Proceedings of the Fourth National Symposium on Sensor Fusion*, Vol. I (Unclassified), Apr. 2-4 1991, Orlando, ERIM, Ann Arbor MI, pp. 173-189

51. A.I. Orlov (1978) "Fuzzy and random sets" (in Russian), *Prikladnoi Mnogomerni Statisticheskii Analys*, Moscow

52. R.K. Pathria (1972) *Statistical Mechanics*, Oxford University Press

53. P. Quinio and T. Matsuyama (1991) "Random Closed Sets: A Unified Approach to the Representation of Imprecion and Uncertainty," in R. Kruse and P. Siegel, eds., *Symbolic and Quantitative Approaches to Uncertainty*, Springer-Verlag, pp. 282-286

54. L.E. Rasmussen (1994) "Approximating the Permanent: A Simple Approach," *Random Structures and Algorithms*, vol. 5 no. 2, pp. 349-361

55. T.L. Saaty and J. Bram (1964) *Nonlinear Mathematics*, Dover Publications

56. G.E. Shilov and B.L. Gurevich, *Integral, Measure, and Derivative: A Unified Approach*, Prentice-Hall, 1966

57. P. Smets (1992) "The Transferable Belief Model and Random Sets," *International Journal of Intelligent Systems*, Vol. 7, pp. 37-46

58. L. Tierney and J.B. Kadane (1986) "Accurate Approximations for Posterior Moments and Marginal Densities," *J. Amer. Stat. Ass.*, vol. 81 no. 392, pp. 82-86

59. R.B. Washburn (1987) "A random point process approach to multiobject tracking," *Proc. American Control Conf.*, vol. 3, pp. 1846-1852

60. E. Waltz and J. Llinas, *Multisensor Data Fusion*, Artech House, Boston 1990

61. R.L. Wheedon and A. Zygmund (1977) *Measure and Integral: An Introduction to Real Analysis*, Marcel Dekker, Inc.

62. H.W. Sorenson and D.L. Alspach (1971) "Recursive Bayesian Estimation Using Gaussian Sums," *Automatica*, vol. 7, pp. 465-479

63. D.L. Alspach (1975) "A Gaussian Sum Approach to the Multi-Target Identification-Tracking Problem," *Automatica*, vol. 11, pp. 285-296

64. J. T.-H. Lo (1994) "Synthetic Approach to Optimal Filtering," *IEEE Trans. Neural Networks*, vol. 5 no. 5, pp. 803-811

65. N. Portenko, H. Salehi, and A. Skorokhod (1997) "On optimal filtering of multitarget tracking systems based on point processes observations," *Random Operators and Stochastic Equations*, vol. 1, pp. 1-34